

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	Time Stamp
L1	10	MODEL AND (INFLUENC\$3 WITH OUTCOME) AND RELATIONSHIP AND METADATA	US-PGPUB; USPAT; EPO; IBM_TDB	OR	ON	2005/12/20 14:24
L2	1	1 AND @AD<"20010330"	US-PGPUB; USPAT; EPO; IBM_TDB	OR	ON	2005/12/20 14:23
L3	128	MODEL AND (INFLUENC\$3 AND OUTCOME) AND RELATIONSHIP AND METADATA	US-PGPUB; USPAT; EPO; IBM_TDB	OR	ON	2005/12/20 14:24
L4	48	3 AND @AD<"20010330"	US-PGPUB; USPAT; EPO; IBM_TDB	OR	ON	2005/12/20 14:24
L5	42	4 AND CONDITION AND QUERY	US-PGPUB; USPAT; EPO; IBM_TDB	OR	ON	2005/12/20 14:25
L6	3	5 AND (METADATA WITH MODEL)	US-PGPUB; USPAT; EPO; IBM_TDB	OR	ON	2005/12/20 14:28
L7	1	5 AND (ADAPTIVE WITH PATTERN)	US-PGPUB; USPAT; EPO; IBM_TDB	OR	ON	2005/12/20 14:31
L8	1	5 AND (ADAPTIVE WITH MODEL)	US-PGPUB; USPAT; EPO; IBM_TDB	OR	ON	2005/12/20 14:36
L9	2	5 AND (PREDICTIVE WITH MODEL)	US-PGPUB; USPAT; EPO; IBM_TDB	OR	ON	2005/12/20 14:31
L10	1	5 AND (HYPOTHESIS)	US-PGPUB; USPAT; EPO; IBM_TDB	OR	ON	2005/12/20 14:36
L11	3	4 AND (HYPOTHESIS)	US-PGPUB; USPAT; EPO; IBM_TDB	OR	ON	2005/12/20 14:40
L12	0	4 AND (HYPOTHESIS WITH MODEL)	US-PGPUB; USPAT; EPO; IBM_TDB	OR	ON	2005/12/20 14:40

L13	3	3 AND (HYPOTHESIS WITH MODEL)	US-PGPUB; USPAT; EPO; IBM_TDB	OR	ON	2005/12/20 14:40
-----	---	----------------------------------	--	----	----	------------------

59/3,K/166 (Item 74 from file: 349)  
DIALOG(R)File 349:PCT FULLTEXT  
(c) 2005 WIPO/Univentio. All rts. reserv.

00141500

**EXPERT SYSTEM USING PATTERN RECOGNITION TECHNIQUES**  
**SYSTEME EXPERT UTILISANT DES TECHNIQUES DE RECONNAISSANCE DE STRUCTURES**

Patent Applicant/Assignee:

HEWLETT PACKARD LIMITED,  
CRAWFORD Thomas Maxton,  
MARTON Virgil Nicolae,  
COWAN Colin Frederick Nathaniel,

Inventor(s):

CRAWFORD Thomas Maxton,  
MARTON Virgil Nicolae,  
COWAN Colin Frederick Nathaniel,

Patent and Priority Information (Country, Number, Date):

Patent: WO 8706371 A1 19871022  
Application: WO 86GB199 19860410 (PCT/WO GB8600199)  
Priority Application: WO 86GB199 19860410

Designated States:

(Protection type is "patent" unless otherwise stated - for applications prior to 2004)

JP US

Publication Language: English

Fulltext Word Count: 3888

**EXPERT SYSTEM USING PATTERN RECOGNITION TECHNIQUES**

Main International Patent Class: **G06F-015/18**

International Patent Class: **G06F-11:22 ...**

**... G06F-15:20**

Fulltext Availability:

Detailed Description

Claims

English Abstract

...pattern recognition facility (65) which can be trained or instructed to recognise different combinations of **input** data as indicative of particular **conditions** of the equipment. The system is trained by coupling it initially to known good equipment into which definable **faults** are introduced. The adaptive **pattern** recognition facility (65) has a series of sections with adjustable weight coefficients which are adjusted...

Detailed Description

**EXPERT SYSTEM USING PATTERN RECOGNITION TECHNIQUES**  
**DESCRIPTION**

Technical Field

This invention relates to expert systems, Such systems can be @sed for monitoring and **fault** diagnosis in a wide variety of applications.

Prior Art

Conventional expert systems are usually operated...

...consuming and costly process. Furthermore it some.

times involves a user predicting the implications of **faults** in a **theoretical** way and it is doubtful whether this is practical as it involves design staff in...

...combinations as representative of particular conditions whereby after instruction the processing means can operate on **input** data to provide an indication of the **condition** of equipment or systems under diagnosis.

The **input** data can be fed to the processing means from the equipment or system under test...adaptive pattern recognition facility

which can be trained or instructed to recognise different combinations of **input** data as indicative of particular **conditions** of the equipment. The adaptive **pattern** recognition facility includes an adaptive combiner in which adjustable coefficients or weights are combined with...

...which provides the input data. Data is input from the equipment 11 so that the **system** 10 initially **recognises** the data **pattern** for that **fault** free equipment. Appropriate coefficients for a section of the adaptive combiner are generated and stored...the fault can be provided.

In summary therefore one combiner section is created for each **outcome** ( **Fault** ) whose presence needs to be detected and measured. The combiner weights reflect the amount of...

...be used to diagnose a wide variety of systems or items of equipment in which **pattern recognition techniques** can be used to identify **faults** . The example to be described is ...centered on the centre of the constellation display.

- 12

It is therefore possible to evaluate **statistical data** relating to the constellation samples from the basic digitised X and Y values of the...when the initial classification indicates that, although not fault free, the radio is reasonably well **conditioned** and any **faults** are such that they can be analysed by the block 65. Broadly speaking block 65...

...is instructed to recognise certain patterns of input data and to associate these with known **fault conditions** so that when diagnosing a digital radio it can produce output signals representative of that **fault condition** . These output signals can be applied to for example a visual display unit to provide... ,

...W is the classifier weight vector for outcome i.

i

- 17

Thus for a particular **condition** ' an **error** value is generated as follows.

e Y Y

Each combiner section is then updated using...

#### Claim

... combinations as representative of particular conditions whereby after instruction the processing means can operate on **input** data to provide an indication of the **condition** of equipment or systems under diagnosis.

2 An expert system as claimed in claim 1...

Set	Items	Description
S1	9588546	IDENTIF? OR LOCAT? OR NARROW?()DOWN? OR FIND? OR RETRIEV? - OR TRACK? OR REVEAL? OR ISOLAT?
S2	2611397	ASCERTAIN? OR DISCERN? OR SIFT? OR FILTER? OR RECOGN? OR D- ISTINGUISH? OR UNCOVER?
S3	4248659	SINGL?()OUT OR PINPOINT? OR SORT??? OR INDICAT? OR DESIGNA- T? OR TURN?()UP OR UNMASK? OR BACKPROPAGAT?
S4	10376182	DETECT? OR DISCOVER? OR UNEARTH? OR EXPOSE? OR EXPOSING? OR CULL? OR FERRET? OR DETERMIN?
S5	4488035	S1:S4(7N)(METHOD? OR SYSTEM? OR PROCESS?? OR PROCEDUR? OR - TECHNIQUE? OR MODE? ?)
S6	7070624	PATTERN? OR SIMILAR? OR LIKENESS? OR RELATIONSHIP? OR CONN- ECTION? OR CONCATENAT?
S7	3865184	KINSHIP? OR LINK? OR AFFILIAT? OR CORRESPOND? OR MATCH?
S8	4897156	ASSOCIATION? OR CORRELAT? OR COINCIDEN? OR TREND? OR TRAJE- CTOR? OR COINCID?
S9	4815865	INFLUEN? OR IMPACT? OR SWAY? OR LEVERAG? OR BIAS? OR SLANT? OR PROPENSIT?
S10	62604	AFFECTATION? OR IMPING? OR CROSS()POLLEN? OR TAINT? OR SCR- EW?()UP
S11	1903769	ERROR? OR FAULT? OR GLITCH? OR ADULTERAT? OR WARP? OR SPOIL- L? OR BLEEDTHRU? OR BLEEDTHROUGH? OR BLEED?() (THRU OR THROUGH- ?)
S12	1259362	SKEW? OR CONFOUND? OR MONKEY()WRENCH? OR PERTURB? OR MANIP- ULAT? OR CORRUPT?
S13	640752	UNFORESEEN? OR UNKNOWN? OR UNPREDICTAB? OR UNACCOUNT? OR U- NEXPECT? OR UNANTICIPAT? OR OVERLOOK? OR INADVERTENT?
S14	15233345	OUTCOME? OR RESULT? OR CONCLUSION? OR ENDRESULT? OR EVENTU- ALIT? OR CAUSE?(2W)EFFECT?
S15	320184	DESTINAT? OR FINALE? OR COMPLETION? OR CESSATION? OR ENDIN- G? OR PROXIMAT?()CAUSE?
S16	142728	TERMINATION? OR AFTERMATH? OR UPSHOT? OR FALLOUT?
S17	111987	REPERCUSSION? OR CULMINAT? OR RAMIFICAT? OR HISTOGRA?
S18	8992767	SET OR SETS OR GROUP? OR CLUSTER? OR ARRAY? OR ASSEMBL?
S19	5422114	PLURAL? OR SEVERAL? OR MULTIP? OR MULTIT?
S20	2893799	*deleted* COLLECTION? OR ASSORTMENT? OR NUMEROUS? OR SERI- ES? OR NODE?
S21	10955858	*deleted* CONDITION? OR HYPOTHE? OR THEORET? OR PRETEND? - OR SIMULAT?
S22	380035	S5(10N)S6:S8
S23	1509503	(STATIST? OR PROBABILIT? OR PREDICT?) (5N) (ANALY? OR DATA? - OR THEOR? OR FORMULA?)
S24	25903	S22 AND S23
S25	380035	S22 OR S24
S26	7886	S25 AND S6:S8(10N)S9:S13 AND (S6:S13)(10N)S14:S17
S27	834226	(S18:S19 OR COLLECTION? OR ASSORTMENT? OR NUMEROUS? OR SER- IES? OR NODE?) (10N) (CONDITION? OR HYPOTHE? OR THEORET? OR PRE- TEND? OR SIMULAT?)
S28	840	S26 AND S23
S29	47	S28 AND S27
S30	18	S28 AND (S18:S19 OR COLLECTION? OR ASSORTMENT? OR NUMEROUS? OR SERIES? OR NODE?) (10N) (ENVIRONMENT? OR SITUATION? OR CONT- INGEN? OR QUALIFICATION?)
S31	3	S28 AND (S18:S19 OR COLLECTION? OR ASSORTMENT? OR NUMEROUS? OR SERIES? OR NODE?) (10N) (STIPULATION? OR EVENT? OR CIRCUMST- AN? OR HAPPENING?)
S32	29	S28 AND (S18:S19 OR COLLECTION? OR ASSORTMENT? OR NUMEROUS? OR SERIES? OR NODE?) (10N) (PHENOMEN? OR INCIDENT? OR OCCURREN- C? OR EPISODE? OR SCENARIO? OR VECTOR?)
S33	401	S28 AND (MODEL? ? OR (TEST OR INSTANT OR REFERENCE OR IDEA- L?) () (CASE? OR SET OR SETS) OR PROFIL?)

S34 208 S33 AND (S18:S19 OR COLLECTION? OR ASSORTMENT? OR NUMEROUS?  
 OR SERIES? OR NODE?)  
 S35 86 S29:S32  
 S36 93 S34 AND (S18:S19 OR COLLECTION? OR ASSORTMENT? OR NUMEROUS?  
 OR SERIES? OR NODE?) (10N) (MODEL? OR (TEST OR INSTANT OR REFE-  
 RENCE OR IDEAL?) () (CASE? OR SET OR SETS))  
 S37 7 S34 AND (S18:S19 OR COLLECTION? OR ASSORTMENT? OR NUMEROUS?  
 OR SERIES? OR NODE?) (10N) PROFIL?  
 S38 165 S35:S37  
 S39 99 S38 AND PY<2002  
 S40 77 RD (unique items)  
 File 2:INSPEC 1969-2005/Jul W5  
 (c) 2005 Institution of Electrical Engineers  
 File 6:NTIS 1964-2005/Jul W5  
 (c) 2005 NTIS, Intl Cpyrght All Rights Res  
 File 8:Ei Compendex(R) 1970-2005/Jul W5  
 (c) 2005 Elsevier Eng. Info. Inc.  
 File 34:SciSearch(R) Cited Ref Sci 1990-2005/Jul W5  
 (c) 2005 Inst for Sci Info  
 File 35:Dissertation Abs Online 1861-2005/Jul  
 (c) 2005 ProQuest Info&Learning  
 File 62:SPIN(R) 1975-2005/May W5  
 (c) 2005 American Institute of Physics  
 File 65:Inside Conferences 1993-2005/Aug W1  
 (c) 2005 BLDSC all rts. reserv.  
 File 94:JICST-EPlus 1985-2005/Jun W3  
 (c) 2005 Japan Science and Tech Corp(JST)  
 File 95:TEME-Technology & Management 1989-2005/Jul W1  
 (c) 2005 FIZ TECHNIK  
 File 99:Wilson Appl. Sci & Tech Abs 1983-2005/Jul  
 (c) 2005 The HW Wilson Co.  
 File 111:TGG Natl.Newspaper Index(SM) 1979-2005/Aug 10  
 (c) 2005 The Gale Group  
 File 144:Pascal 1973-2005/Jul W5  
 (c) 2005 INIST/CNRS  
 File 239:Mathsci 1940-2005/Sep  
 (c) 2005 American Mathematical Society  
 File 256:TecInfoSource 82-2005/Jul  
 (c) 2005 Info.Sources Inc  
 File 434:SciSearch(R) Cited Ref Sci 1974-1989/Dec  
 (c) 1998 Inst for Sci Info  
 ?

40/3,K/8 (Item 8 from file: 2)

DIALOG(R)File 2:INSPEC

(c) 2005 Institution of Electrical Engineers. All rts. reserv.

5650023 INSPEC Abstract Number: B9709-6130-047, C9709-1250C-014

**Title: A statistical language model for speech recognition**

Author(s): O'Boyle, P.

Author Affiliation: Dept. of Comput. Sci., Queens Univ., Belfast, UK

Conference Title: Proceedings of the IASTED International Conference  
Artificial Intelligence, Expert Systems and Neural Networks p.52-5

Editor(s): Hamza, M.H.

Publisher: IASTED, Anaheim, CA, USA

Publication Date: 1994 Country of Publication: USA 178 pp.

ISBN: 0 88986 200 1 Material Identity Number: XX95-01476

Conference Title: Proceedings IASTED/ISMM Symposium. Artificial  
Intelligence, Expert Systems and Neural Networks

Conference Sponsor: IASTED

Conference Date: 4-6 July 1994 Conference Location: Zurich,  
Switzerland

Language: English

Subfile: B C

Copyright 1997, IEE

**Title: A statistical language model for speech recognition**

Abstract: A new statistical language model, which is suitable for use in natural language speech recognition, has been developed. The model combines maximum likelihood probabilities derived from the analysis of a text corpus to produce conditional probabilities for words following sequences of words. The model is therefore similar to the trigram models developed by the IBM continuous speech recognition group. However, a new approach is taken when combining the probability distributions. This allows the model to be set up more rapidly and permits quicker modification of the model's training corpus. The model has been incorporated as part of a system to reconstruct sentences from corrupted phoneme sequences. This presents a task similar to that faced by language models in automatic speech recognition. Since the corrupted sequences were generated using a confusion matrix for phonemes derived from actual speech the task is also of similar difficulty to speech recognition. By comparing the results for the system with the language model with those produced when the model is removed we find that in our experiments the model reduces the error rate by a factor of 10 and more than doubles the number...

Identifiers: statistical language model ; ...

...trigram models ; ...

...IBM continuous speech recognition group ;  
1994



40/3,K/11 (Item 11 from file: 2)

DIALOG(R)File 2:INSPEC

(c) 2005 Institution of Electrical Engineers. All rts. reserv.

04074831 INSPEC Abstract Number: C9203-0100-002

**Title: Statistical mechanics of combat with human factors**

Author(s): Ingber, L.; Swarder, D.D.

Author Affiliation: Sci. Transfer Corp., McLean, VA, USA

Journal: Mathematical and Computer Modelling vol.15, no.11 p.99-127

Publication Date: 1991 Country of Publication: UK

CODEN: MCMOEG ISSN: 0895-7177

U.S. Copyright Clearance Center Code: 0895-7177/91/\$3.00+0.00

Language: English

Subfile: C

Abstract: A combat model based on **statistical** mechanics is extended using control **theory** to include human factors. Similar scenarios differing at crucial decision points are used for data to model the **influence** of decision-making on combat. The **results** may then be used to improved human factors and C/sup 2/ algorithms in computer...

...representation is used, more suitable for numerical and formal work, i.e. a Lagrangian representation. **Theoretically**, these equations are nested within a larger **set** of nonlinear stochastic operator-equations which include C/sup 3/ human factors, e.g. supervisory decisions. The authors perturb this operator theory about the SMC zeroth order **set** of equations. Then, subsets of **scenarios** fit to zeroth order, originally considered to be **similar** degenerate, can be farther split **perturbatively** to **distinguish** C/sup 3/ decision-making influences. **Methods** of very fast simulated re-annealing are used for fitting these models to empirical data.

...Descriptors: **statistical analysis** ;

1991

40/3,K/17 (Item 4 from file: 6)

DIALOG(R) File 6:NTIS

(c) 2005 NTIS, Intl Cpyrght All Rights Res. All rts. reserv.

1361426 NTIS Accession Number: AD-A189 070/6

**Detecting Serial Correlation in the Error Structure of a Cross-Lagged Panel Model**

(Technical rept)

Mayer, L. S. ; Carroll, S. S.  
Stanford Univ., CA. Dept. of Statistics.

Corp. Source Codes: 009225032; 332580

Report No.: TR-401

2 Feb 88 25p

Languages: English

Journal Announcement: GRAI8812

Order this product from NTIS by: phone at 1-800-553-NTIS (U.S. customers); (703)605-6000 (other countries); fax at (703)321-8547; and email at orders@ntis.fedworld.gov. NTIS is located at 5285 Port Royal Road, Springfield, VA, 22161, USA.

NTIS Prices: PC A03/MF A01

**Detecting Serial Correlation in the Error Structure of a Cross-Lagged Panel Model**

... two or more variable are measured for a large number of subjects at each of **several** waves or points in time. The variables divide naturally into two **sets** and the primary purpose of analysis is to estimate and test the strength of the relationship between the **sets**. This paper contributes to these studies by developing and applying **procedures** for **detecting** the presence of serial **correlation** in the **error** structure of the regression **models** used in such studies. The regression approach was extended by incorporating the cross-effects as parameters in a multivariate regression **model** and develops procedures to estimate and test these parameters. Both the **model** with independent **errors** and the **model** with serially **correlated errors** were considered. This paper extends the applicability of this **results** by considering the problem of serially **correlated error** structure as opposed to the independent **error** structure.

Descriptors: \*Regression analysis; \*Time **series** analysis; Errors; Estimates; Mathematical **models**; Multivariate **analysis**; Statistical **analysis**

40/3,K/18 (Item 5 from file: 6)

DIALOG(R)File 6:NTIS

(c) 2005 NTIS, Intl Cpyrght All Rights Res. All rts. reserv.

1243569 NTIS Accession Number: AD-A166 477/0

**Power of Statistical Tests Used in Correlation Techniques for  
Battlefield Identification**

(Final rept)

Myhre, J. ; Rennie, M. ; Duquette, W. ; Hockman, D.

Jet Propulsion Lab., Pasadena, CA.

Corp. Source Codes: 014828000; 191150

Report No.: JPL-D-2793

Aug 85 137p

Languages: English

Journal Announcement: GRAI8616

Prepared in cooperation with Claremont McKenna Coll., Claremont, CA.  
Inst. of Decision Science.

Order this product from NTIS by: phone at 1-800-553-NTIS (U.S. customers); (703)605-6000 (other countries); fax at (703)321-8547; and email at orders@ntis.fedworld.gov. NTIS is located at 5285 Port Royal Road, Springfield, VA, 22161, USA.

NTIS Prices: PC A07/MF A01

**Power of Statistical Tests Used in Correlation Techniques for  
Battlefield Identification**

This report is the second in a series presenting a study of self-**correlation** algorithms in intelligence **systems**. These algorithms use multivariate statistical tests to **determine** the equality of mean **vectors** from two different data **sets**. For example, tests are used to determine the equality of location **vectors** from two different data **sets** (Are the data from the same emitter). This report considers estimation of the probability that...

... frequency of test errors were estimated by simulation for most of the cases studied. The **results** indicate that in some of the cases the **error** rate is high enough to be of possible concern. Keywords: Robustness; Statistical distributions.

Descriptors: \***Correlatio** n **techniques** ; \* **Statistical** tests; Multivariate **analysis** ; Battlefields; **Identification** ; **Data bases** ; **Errors** ; Rates; Estimates; Position(Location); Frequency; Mean; **Data** management; Distribution; Simulation; **Probability** ; Intelligence; **Statistical** distributions; Skewness

40/3,K/19 (Item 6 from file: 6)

DIALOG(R)File 6:NTIS

(c) 2005 NTIS, Intl Copyright All Rights Res. All rts. reserv.

1139642 NTIS Accession Number: ORNL/CSD/TM-65

**Probabilistic Method for Grouping Data**

Begovich, C. L.

Union Carbide Corp., Oak Ridge, TN. Nuclear Div.

Corp. Source Codes: 004624008; 6433000

Sponsor: Department of Energy.

Nov 79 191p

Languages: English

Journal Announcement: GRAI8426; NSA0500

Order this product from NTIS by: phone at 1-800-553-NTIS (U.S. customers); (703)605-6000 (other countries); fax at (703)321-8547; and email at orders@ntis.fedworld.gov. NTIS is located at 5285 Port Royal Road, Springfield, VA, 22161, USA.

NTIS Prices: PC A09/MF A01

... cluster analysis procedures. In the analysis, the data are perturbed by Monte Carlo techniques to **simulate** the experimental **error**, and the **resultant** data **sets** are **clustered**. By varying the number of clusters, a procedure is given to estimate the unknown number...

...different examples. The probabilistic method is shown to have advantages for determining the number of **groups** and the probabilities for a sample's membership in the **hypothesized groups**. 33 figures, 8 tables. (ERA citation 05:006396)

Descriptors: \*Data analysis; \* **Pattern recognition** ; Algorithms; Fortran; Ibm computers; Monte carlo **method** ; P codes; **Perturbation theory** ; **Probability** ; **Statistics**

40/3,K/21 (Item 8 from file: 6)

DIALOG(R)File 6:NTIS

(c) 2005 NTIS, Intl Cpyrght All Rights Res. All rts. reserv.

0612913 NTIS Accession Number: AD-813 395/1/XAB

**Data Reduction Using Information Theoretic Techniques**

(Final rept. Jan-Dec 66)

Hughes, G. F. ; Lebo, J. A.

North American Aviation Inc Anaheim Calif Autonetics Div

Corp. Source Codes: 401838

Report No.: C7-147/501; RADC-TR-67-67

Mar 67 155p

Journal Announcement: GRAI7710

Distribution limitation now removed. Order this product from NTIS by: phone at 1-800-553-NTIS (U.S. customers); (703)605-6000 (other countries); fax at (703)321-8547; and email at orders@ntis.fedworld.gov. NTIS is located at 5285 Port Royal Road, Springfield, VA, 22161, USA.

NTIS Prices: PC A08/MF A01

Final results are presented from a mathematical study to improve the overall performance of **pattern** recognition devices, such as for reconnaissance photograph interpretation. The specific results of this study are: (1) the design of practical **pattern recognition systems** requires mathematical **procedures** to account for statistical interdependencies and redundancy. (2) Inadequate experimental data **sets** are a common **situation** in practice and quantitative error formulas from sampling theory should be employed to assess and...

... redundancy and data inadequacy. (4) A method of evaluating competing recognizer designs was developed from **statistical sampling theory**.

Descriptors: **\*Patter** n recognition; Mathematical analysis; Photointerpretation; **Errors** ; Sampling; Equations; Computer programs; Digital computers; Experimental data; **Set** theory; **Data** processing; Functional **analysis** ; Mathematical **models** ; **Probability**

40/3,K/28 (Item 6 from file: 8)  
DIALOG(R)File 8:EI Compendex(R)  
(c) 2005 Elsevier Eng. Info. Inc. All rts. reserv.

05111545 E.I. No: EIP98094366286

**Title:** Statistical analysis with bootstrap diagnostics of atmospheric pollutants predicted in the APSIS experiment

Author: Archer, Graeme; Giovannoni, Jean-Michel

Corporate Source: Environmental Inst, Ispira (VA), Italy

Source: Water, Air and Soil Pollution v 106 n 1-2 Aug 1998. p 43-81

Publication Year: 1998

CODEN: WAPLAC ISSN: 0049-6979

Language: English

**Title:** Statistical analysis with bootstrap diagnostics of atmospheric pollutants predicted in the APSIS experiment

**Abstract:** Predictions of O<sub>3</sub> and NO<sub>2</sub> delivered by three eulerian models from the APSIS study are compared with observed levels of the phenomena recorded on 25...

...1990, over the greater Athens area, Greece. A variety of measures are used to test model performances, including normalized mean square error, bias and correlation, for which confidence intervals are constructed. Causes of differences in the model differences are briefly discussed. A jackknife diagnostic check is performed on the adequacy of the bootstrap procedure used to make the intervals, which are found to be suitably robust. Similarities in model behaviour, and specific performances, are detected. Some statistical criteria seem to give a slight advantage to one of the three sets of model results, although important biases can occur locally, between observations and predictions. On the whole, no one of the set of model results emerges as significantly superior to the others. (Author abstract) 16 Refs.

**Descriptors:** \*Air pollution; Ozone; Nitrogen oxides; Mathematical models; Photochemical reactions; Statistical methods; Atmospheric composition; Error analysis; Correlation methods

**Identifiers:** Eulerian models; Bootstrap diagnostics; Jackknife diagnostics

40/3,K/43 (Item 7 from file: 34)  
DIALOG(R)File 34:SciSearch(R) Cited Ref Sci  
(c) 2005 Inst for Sci Info. All rts. reserv.

05286245 Genuine Article#: VM998 No. References: 30

Title: **LINEAR-REGRESSION FOR ASTRONOMICAL DATA WITH MEASUREMENT ERRORS AND  
INTRINSIC SCATTER**

Author(s): AKRITAS MG; BERSHADY MA

Corporate Source: PENN STATE UNIV,DEPT STAT/STATE COLL//PA/16802; PENN  
STATE UNIV,DEPT ASTRON & ASTROPHYS/STATE COLL//PA/16802

Journal: ASTROPHYSICAL JOURNAL, 1996 , V470, N2 (OCT 20), P706-714

ISSN: 0004-637X

Language: ENGLISH Document Type: ARTICLE (Abstract Available)

, 1996

...Abstract: orthogonal regression can be constructed. We refer to this method as BCES estimators (for bivariate **correlated errors** and intrinsic scatter). The second method is a weighted least squares (WLS) estimator, which applies...

...held galaxies, yields significantly different slopes than OLS and other estimators used in the literature. **Simulations** with artificial data **sets** are used to evaluate the small sample performance of the estimators. Not surprisingly, the least- **biased results** are obtained when color is treated as the dependent variable. The Tully-Fisher relation is another example for which the BCES method should be used because **errors** in luminosity and velocity are **correlated** because of inclination corrections. We also **find**, via simulations, that the WLS **method** is by far the best method for the Tolman surface-brightness test, producing the smallest...

Research Fronts: 94-0945 001 ( **ERRORS** -IN-VARIABLES MODELS; URINARY ELECTROLYTE EXCRETION; LINEAR FUNCTIONAL- **RELATIONSHIPS** ; RANDOM NUMBER)

40/3,K/44 (Item 8 from file: 34)  
DIALOG(R)File 34:SciSearch(R) Cited Ref Sci  
(c) 2005 Inst for Sci Info. All rts. reserv.

05012729 Genuine Article#: UZ125 No. References: 43  
**Title:** CLUSTER CORRELATION-FUNCTIONS IN N-BODY SIMULATIONS  
**Author(s):** EKE VR; COLE S; FRENK CS; NAVARRO JF  
**Corporate Source:** UNIV DURHAM, SCI LABS, DEPT PHYS, SOUTH RD/DURHAM DH1  
3LE//ENGLAND/; UNIV ARIZONA, STEWARD OBSERV/TUCSON//AZ/85721  
**Journal:** MONTHLY NOTICES OF THE ROYAL ASTRONOMICAL SOCIETY, 1996 , V281,  
N2 (JUL 15), P703-715  
**ISSN:** 0035-8711  
**Language:** ENGLISH **Document Type:** ARTICLE (Abstract Available)

**Title:** CLUSTER CORRELATION-FUNCTIONS IN N-BODY SIMULATIONS  
, 1996

**Abstract:** The correlation function of galaxy clusters has frequently been used as a test of cosmological models. A number of assumptions are implicit in the comparison of theoretical expectations with data. Here we use an ensemble of ten large N-body simulations of the standard cold dark matter cosmology to investigate how cluster selection criteria and other uncertain factors influence the cluster correlation function. Our study is restricted to the idealized case where clusters are identified in the three-dimensional mass distribution of the simulations. We consider the effects of varying the definition of a cluster, the mean number density (or equivalently the threshold richness or luminosity) in a catalogue, and the assumed normalization of the cosmological model; we also examine the importance of redshift space distortions. We implement five different group-finding algorithms and construct cluster catalogues defined by mass, velocity dispersion or a measure of X-ray luminosity. We find that different cluster catalogues yield correlation functions which can differ from one another by substantially more than the statistical errors in any one determination. For example, at a fixed number density of clusters, the characteristic clustering length can vary by up to a factor of similar to 1.5, depending on the precise procedure employed to identify and select clusters. For a given cluster selection criterion, the correlation length typically varies by similar to 20 per cent in catalogues spanning the range of...

...in redshift space enhance the correlation function at large separations and lead to a larger clustering length in redshift space than in real space. The sensitivity of the cluster correlation function to various uncertain model assumptions substantially weakens previous conclusions based on the comparison of model predictions with real data. For example, some of our standard cold dark matter cluster catalogues agree better with published cluster clustering data (particularly on small and intermediate scales) than catalogues constructed from similar simulations by Bahcall & Cen and Croft & Efstathiou. Detailed modelling of cluster selection procedures including, for example, the effects of selecting from projected galaxy catalogues is required before the cluster correlation function can be regarded as a high-precision constraint on cosmological models.

...Identifiers--X-RAY- CLUSTERS ; COLD DARK MATTER; SPATIAL CORRELATION-FUNCTION; SOUTHERN GALAXY CATALOG; RICH CLUSTERS ; ABELL CLUSTERS ; EXPANDING UNIVERSE; REDSHIFT SURVEY; APM; EVOLUTION  
...Research Fronts: 005 (COSMIC MICROWAVE BACKGROUND ANISOTROPIES; 1.2-JY IRAS GALAXY REDSHIFT SURVEY; DOUBLE INFLATIONARY COSMOLOGICAL MODEL ; PRIMORDIAL POWER SPECTRUM)



94-0059 002 (RICH **CLUSTERS** OF GALAXIES; WARM ABSORBER IN ABELL 2256;  
RADIO DATA)

40/3,K/48 (Item 12 from file: 34)  
DIALOG(R)File 34:SciSearch(R) Cited Ref Sci  
(c) 2005 Inst for Sci Info. All rts. reserv.

02740339 Genuine Article#: LZ807 No. References: 19  
**Title: GENERALIZED CLUSTERING NETWORKS AND KOHONEN SELF-ORGANIZING SCHEME**  
Author(s): PAL NR; BEZDEK JC; TSAO ECK  
Corporate Source: UNIV W FLORIDA, DEPT COMP SCI, DIV COMP  
SCI/PENSACOLA//FL/32514; UNIV CALCUTTA/CALCUTTA 700073/W BENGAL/INDIA/;  
INDIAN STAT INST, ELECTR & COMMUN SCI UNIT/CALCUTTA 700035/W  
BENGAL/INDIA/  
Journal: IEEE TRANSACTIONS ON NEURAL NETWORKS, 1993, V4, N4 (JUL), P  
549-557  
ISSN: 1045-9227  
Language: ENGLISH Document Type: ARTICLE (Abstract Available)

**Title: GENERALIZED CLUSTERING NETWORKS AND KOHONEN SELF-ORGANIZING SCHEME**  
**, 1993**

Abstract: This paper first discusses the relationship between the sequential hard c-means (SHCM) and learning **vector** quantization (LVQ) **clustering** algorithms. These methods suffer from **several** major problems. For example, they depend heavily on initialization. If the initial values of the **cluster** centers are outside the convex hull of the input data, such algorithms, even if they terminate, may not produce meaningful results in terms of prototypes for **clustering**. This is due in part to the fact that they update only the winning prototype...

...families of methods with Kohonen's self-organizing feature mapping (SOFM), which is not a **clustering** method, but which often lends ideas to **clustering** algorithms. Finally, we propose a generalization of LVQ which (may) update all **nodes** for a given input **vector**. Moreover, our network attempts to find a minimum of a well-defined objective function. The learning rules depend on the degree of distance match to the winner **node**; the lesser the degree of **match** with the winner, the more is the **impact** on nonwinner **nodes**. Numerical **results** indicate that the terminal prototypes generated by this modification of LVQ are generally insensitive to...

Research Fronts: 91-0754 002 (NEURAL NETWORKS; LOCAL LEARNING  
PRESCRIPTION FOR ARBITRARY **CORRELATED PATTERNS**; **STATISTICAL**  
-MECHANICAL **FORMULATION OF THE WILLSHAW MODEL**)  
91-2579 002 (FUZZY CLUSTERING; **PATTERN - RECOGNITION SYSTEM**;  
AUTOMATIC **IDENTIFICATION**; CLASSIFICATION OF IMAGES)  
91-3153 001 (ASYMPTOTICALLY EFFICIENT ADAPTIVE-CONTROL OF LINEAR  
STOCHASTIC-SYSTEMS; PARALLEL RECURSIVE ALGORITHMS; PERFORMANCE GRADIENT  
ESTIMATION; ARMA **MODELS**)

40/3,K/63 (Item 13 from file: 35)  
DIALOG(R)File 35:Dissertation Abs Online  
(c) 2005 ProQuest Info&Learning. All rts. reserv.

01142618 ORDER NO: AADDX-91341

**CHARACTERIZATION OF THE ASSOCIATION BETWEEN SHORT TERM VARIATIONS IN DAILY  
MORTALITY AND ADVERSE ENVIRONMENTAL CONDITIONS USING TIME SERIES  
METHODOLOGY**

Author: GUZMAN, MARTHA ELVA RAMIREZ  
Degree: PH.D.  
Year: 1990  
Corporate Source/Institution: UNIVERSITY OF READING (UNITED KINGDOM) (0354)  
Source: VOLUME 51/09-B OF DISSERTATION ABSTRACTS INTERNATIONAL.  
PAGE 4430. 284 PAGES

**CHARACTERIZATION OF THE ASSOCIATION BETWEEN SHORT TERM VARIATIONS IN DAILY  
MORTALITY AND ADVERSE ENVIRONMENTAL CONDITIONS USING TIME SERIES  
METHODOLOGY**

Year: 1990

...association with The British Library.

Most of the literature review reports the use of regression **models** to investigate mortality-environmental relationships. The goal of this research work has been to devise improved methods of **statistical** and computational **analyses** to study mortality associated with daily weather and pollution concentrations making allowance for influenza epidemics...

...environmental variables has been (1) To describe the main characteristics of mortality, pollution and weather **series** with univariate ARIMA **models**. (2) To distinguish between sporadic sudden changes and step level changes increases in mortality. (3) To assess excess mortality associated with influenza. (4) To prewhiten all mortality, weather and pollution **series** to allow for **influenza** epidemics in the mortality- **environmental modelling relationship**. (5) To build up transfer function **models** to capture the variability of both the mortality and the environmental variables. (6) To develop a methodology for building a common **model** to describe a **cause** -effect time series **relationship** when **multiple** time **series** are available. This **method** is essentially a combination of **several identification procedures** for transfer function **models**. (7) To develop a computational system on SAS, to facilitate the analysis of time **series**. For this, an interactive system called the TODAY system, and a **set** of specific-purpose programs were created. The TODAY system was fundamental for building up univariate ARIMA time **series models** and for detecting sporadic and level increases in mortality. The specific-purpose programs, complement the...  
...functions and the CORNER program which produces corner tables utilized to identify parsimonious transfer function **models** ..

The transfer function **models** which related temperature and smoke with mortality, indicates that deaths are related to the cumulative...

40/3,K/64 (Item 14 from file: 35)  
DIALOG(R)File 35:Dissertation Abs Online  
(c) 2005 ProQuest Info&Learning. All rts. reserv.

1053505 ORDER NO: AAD82-23196

**COMBINING THE RESULTS OF CORRELATIONAL STUDIES: THEORETICAL  
CONSIDERATIONS AND PRACTICAL APPROACHES**

Author: APLING, RICHARD NELSON

Degree: ED.D.

Year: 1981

Corporate Source/Institution: HARVARD UNIVERSITY (0084)

Source: VOLUME 43/05-A OF DISSERTATION ABSTRACTS INTERNATIONAL.  
PAGE 1512. 223 PAGES

**COMBINING THE RESULTS OF CORRELATIONAL STUDIES: THEORETICAL  
CONSIDERATIONS AND PRACTICAL APPROACHES**

Year: 1981

...both published and unpublished works with few constraints on what studies are included in the **analysis** and aggregates summary **statistics** from studies to produce a distribution of study statistics for which a mean, medium, standard...

...Most of the work on integrating studies has concentrated on combining experimental and quasi-experimental **results**. However, findings from **correlational** studies can also be combined. Since synthesizing **correlations** has received less attention and warrants more **systematic** investigation, this thesis examines issues in combining **findings** from such studies.

Glass raises several issues in his brief discussion of combining correlations, but...

...for the theoretical underpinnings of correlational research. Since a great deal is known about how **correlations** behave and what **influences** the magnitude of these statistics, it seems plausible to use this knowledge to inform decisions about combining **correlational** results.

This thesis reviews the **statistical theory** of the **correlation** coefficient regarding its sampling distribution and effects of measurement **error**, restriction in range, sample size, and other study characteristics on its estimation. Next, the thesis discusses methods for using what we know about **correlations** to inform decisions on how to combine **correlational results**. These methods include using descriptive **statistics**, frequency distributions, Exploratory **Data Analysis**, **hypothesis** testing, and **multiple** regression to examine the behavior of correlations from many studies.

Following this theoretical discussion, the...

...longitudinal data. The thesis concludes by generalizing some of the problems and lessons from combining **correlational results** to broader issues of synthesizing quantitative findings.

Set	Items	Description
S1	18800693	IDENTIF? OR LOCAT? OR NARROW?()DOWN? OR FIND? OR RETRIEV? - OR TRACK? OR REVEAL? OR ISOLAT?
S2	5457694	ASCERTAIN? OR DISCERN? OR SIFT? OR FILTER? OR RECOGN? OR D- ISTINGUISH? OR UNCOVER?
S3	7612098	SINGL?()OUT OR PINPOINT? OR SORT??? OR INDICAT? OR DESIGNA- T? OR TURN?()UP OR UNMASK? OR BACKPROPAGAT?
S4	7740837	DETECT? OR DISCOVER? OR UNEARTH? OR EXPOSE? OR EXPOSING? OR CULL? OR FERRET? OR DETERMIN?
S5	2613991	S1:S4(7N)(METHOD? OR SYSTEM? OR PROCESS?? OR PROCEDUR? OR - TECHNIQUE? OR MODE? ?)
S6	98294	S5(7N)(PATTERN? OR SIMILAR? OR LIKENESS? OR RELATIONSHIP? - OR CONNECTION? OR CONCATENAT? OR EXPECT?)
S7	79540	S5(7N)(KINSHIP? OR LINK? OR AFFILIAT? OR CORRESPOND? OR MA- TCH? OR SIGNIFICAN? OR LIKELIHOOD?)
S8	34907	S5(7N)(ASSOCIATION? OR CORRELAT? OR COINCIDEN? OR TREND? OR TRAJECTOR? OR COINCID? OR ANTICIPAT?)
S9	4705	S6:S8(7N)(INFLUEN? OR IMPACT? OR SWAY? OR LEVERAG? OR BIAS? OR SLANT? OR PROPENSIT?)
S10	27	S6:S8(7N)(AFFECTATION? OR IMPING? OR CROSS()POLLEN? OR TAI- NT? OR SCREW?()UP OR SNAFU?)
S11	2969	S6:S8(7N)(ERROR? OR FAULT? OR GLITCH? OR ADULTERAT? OR WAR- P? OR SPOIL? OR BLEEDTHRU? OR BLEEDTHROUGH? OR BLEED?() (THRU - OR THROUGH?))
S12	502	S6:S8(7N)(SKEW? OR CONFOUND? OR MONKEY()WRENCH? OR PERTURB? OR MANIPULAT? OR CORRUPT?)
S13	885	S6:S8(7N)(UNFORESEEN? OR UNKNOWN? OR UNPREDICTAB? OR UNACC- OUNT? OR UNEXPECT? OR UNANTICIPAT? OR OVERLOOK? OR INADVERTEN- T?)
S14	8939	S9:S13
S15	6789	S14 AND (OUTCOME? OR RESULT? OR CONCLUSION? OR ENDRESULT? - OR EVENTUALIT? OR CAUSE?(2W)EFFECT?)
S16	1240	S14 AND (DESTINAT? OR FINALE? OR COMPLETION? OR CESSATION? OR ENDING? OR PROXIMAT?()CAUSE?)
S17	334	S14 AND (TERMINATION? OR AFTERMATH? OR UPSHOT? OR FALLOUT?)
S18	356	S14 AND (REPERCUSSION? OR CULMINAT? OR RAMIFICAT? OR HISTO- GRA?)
S19	6903	S15:S18
S20	1210	S19 AND (SET OR SETS OR GROUP? OR CLUSTER? OR ARRAY? OR AS- SEMBL? OR PLURAL? OR SEVERAL? OR MULTIP? OR MULTIT? OR COLLEC- TION? OR ASSORTMENT? OR NUMEROUS? OR SERIES? OR NODE?) (5N) (CO- NDITION? OR HYPOTHE? OR THEORET? OR PRETEND? OR SIMULAT?)
S21	631	S19 AND (SET OR SETS OR GROUP? OR CLUSTER? OR ARRAY? OR AS- SEMBL? OR PLURAL? OR SEVERAL? OR MULTIP? OR MULTIT? OR COLLEC- TION? OR ASSORTMENT? OR NUMEROUS? OR SERIES? OR NODE?) (5N) (EN- VIRONMENT? OR SITUATION? OR CONTINGEN? OR QUALIFICATION?)
S22	360	S19 AND (SET OR SETS OR GROUP? OR CLUSTER? OR ARRAY? OR AS- SEMBL? OR PLURAL? OR SEVERAL? OR MULTIP? OR MULTIT? OR COLLEC- TION? OR ASSORTMENT? OR NUMEROUS? OR SERIES? OR NODE?) (5N) (ST- IPULATION? OR EVENT? OR CIRCUMSTAN? OR HAPPENING?)
S23	377	S19 AND (SET OR SETS OR GROUP? OR CLUSTER? OR ARRAY? OR AS- SEMBL? OR PLURAL? OR SEVERAL? OR MULTIP? OR MULTIT? OR COLLEC- TION? OR ASSORTMENT? OR NUMEROUS? OR SERIES? OR NODE?) (5N) (PH- ENOMEN? OR INCIDENT? OR OCCURENC? OR EPISODE? OR SCENARIO? OR VECTOR?)
S24	1522	S19 AND (SET OR SETS OR GROUP? OR CLUSTER? OR ARRAY? OR AS- SEMBL? OR PLURAL? OR SEVERAL? OR MULTIP? OR MULTIT? OR COLLEC- TION? OR ASSORTMENT? OR NUMEROUS? OR SERIES? OR NODE?) (5N) (MO- DEL? ? OR (TEST OR INSTANT OR REFERENCE OR IDEAL?) () (CASE? OR SET OR SETS)
S25	2687	S20:S24

S26 1969 S25 AND (QUERY? OR QUERIE? OR REQUEST? OR INTERROG? OR INQ-  
 UIR? OR SEARCH? OR RETRIEV? OR INPUT? OR INTERFAC?)  
 S27 1419 S25 AND (STATIST? OR PROBABILIT? OR PREDICT?) (5N) (ANALY? OR  
 DATA? OR THEOR? OR FORMULA? OR ALGORITHM?)  
 S28 1057 S26 AND S27  
 S29 1969 S26 OR S28  
 S30 1463 S29 AND PY<2002  
 S31 15 S30 AND S5/TI  
 S32 4 S30 AND S6:S8/TI  
 S33 0 S30 AND S9:S13/TI  
 S34 32 S30 AND S14:S17/TI  
 S35 225 S30 AND S1:S27/TI  
 S36 225 S31:S35  
 S37 174 RD (unique items)  
 File 9:Business & Industry(R) Jul/1994-2005/Aug 12  
 (c) 2005 The Gale Group  
 File 13:BAMP 2005/Aug W1  
 (c) 2005 The Gale Group  
 File 15:ABI/Inform(R) 1971-2005/Aug 13  
 (c) 2005 ProQuest Info&Learning  
 File 16:Gale Group PROMT(R) 1990-2005/Aug 12  
 (c) 2005 The Gale Group  
 File 20:Dialog Global Reporter 1997-2005/Aug 15  
 (c) 2005 Dialog  
 File 47:Gale Group Magazine DB(TM) 1959-2005/Aug 15  
 (c) 2005 The Gale group  
 File 75:TGG Management Contents(R) 86-2005/Aug W1  
 (c) 2005 The Gale Group  
 File 88:Gale Group Business A.R.T.S. 1976-2005/Aug 12  
 (c) 2005 The Gale Group  
 File 98:General Sci Abs/Full-Text 1984-2004/Dec  
 (c) 2005 The HW Wilson Co.  
 File 141:Readers Guide 1983-2004/Dec  
 (c) 2005 The HW Wilson Co  
 File 148:Gale Group Trade & Industry DB 1976-2005/Aug 15  
 (c)2005 The Gale Group  
 File 160:Gale Group PROMT(R) 1972-1989  
 (c) 1999 The Gale Group

?

37/3,K/2 (Item 2 from file: 13)  
DIALOG(R)File 13:BAMP  
(c) 2005 The Gale Group. All rts. reserv.

00691055 Supplier Number: 25914672 (USE FORMAT 7 OR 9 FOR FULLTEXT)

**Modeling Stock Market Returns: Part 3 of 3 parts**

(Instead of opting for either calculating an intrinsic value from theory,  
or **searching** for stable relationships between market and variables  
that appear to drive it, use combination of the two in forecasting stock  
market returns)

Article Author(s): Harasty, Helene; Roulet, Jacques  
Journal of Portfolio Management, v 26, n 2, p 33-46  
Winter 2000

DOCUMENT TYPE: Journal ISSN: 0095-4918 (United States)

LANGUAGE: English RECORD TYPE: Fulltext; Abstract

WORD COUNT: 2013

(USE FORMAT 7 OR 9 FOR FULLTEXT)

(Instead of opting for either calculating an intrinsic value from theory,  
or **searching** for stable relationships between market and variables  
that appear to drive it, use combination of...  
)

**ABSTRACT:**

...that will be reached at a certain time while the other approach is used  
for **searching** for stable relationships between the market and the factors  
driving it. Under the present value...

...and the 1990s. Within this period, two major periods of overvaluation  
occurred, which in turn, **resulted** in substantial corrections. Article  
includes a table showing the effect of interest rate variations in...

**TEXT:**

**SHORT-RUN DYNAMICS AND FORECASTING**

**Results** of estimation of the second equation of the model for the seven  
largest markets are...

...force is significant in all countries but Italy (not shown).(11) Thus,  
the first important **result** is that we are able to identify Granger error  
correction representations between stock prices and...  
...the variables that influence stock markets: mainly EPS, yields, and  
exchange rates. A consistent macroeconomic **scenario** can be built, or even  
**several scenarios**, to examine the probable trend in the market under  
various assumptions, or the sensitivity of...

...The information coefficient (IC), defined as the correlation between the  
forecast and actual returns, and **results** are presented in Exhibit 7 for  
the naive and consensus cases. In both cases, over...

...and short-term rates. Only  
consensus estimates for exchange rates are introduced in the  
forecast.

**CONCLUSION**

Our two-step **error correction model** enables us to **identify**  
cointegrating **relationships** between the stock markets and earnings per  
share at index level and long-term rates...

37/3,K/25 (Item 15 from file: 15)  
DIALOG(R)File 15:ABI/Inform(R)  
(c) 2005 ProQuest Info&Learning. All rts. reserv.

01907814 05-58806

**Specification search and levels of significance in econometric models**  
Caudill, Steven B; Holcombe, Randall G  
Eastern Economic Journal v25n3 PP: 289-300 Summer 1999  
ISSN: 0094-5056 JRNL CODE: EEJ  
WORD COUNT: 5230

**Specification search and levels of significance in econometric models**

...DESCRIPTORS: **Searches** ;

ABSTRACT: The problem specification **searches** pose for inference are described, the **results** of some simulations are presented for purposes of illustration, and the bootstrapping method is used...

...of statistical significance than a standard t-test. The value of the illustrations of specification **searches** is that they help demonstrate the severity of the problem. The examples presented illustrate that in most cases, a researcher can undertake a specification **search** and report a statistically significant **result** regardless of whether the variables in a regression equation are actually related. The bootstrap procedure used to analyze the specification **searches** does provide another way to examine the true statistical significance of empirical **results**. Two different specification **searches** are examined: a drop insignificant coefficients **search** and a biggest t-ratio **search**. Both are shown to lead to larger than reported standard errors. ...

...TEXT: and frequently they are.1 Often, empirical articles in economic journals display tables of regression **results** showing the effects of adding or deleting variables.2 Also common is the reporting of **results** in linear and log linear form, and the examination of trend stationary versus difference stationary...

...even though no mention of the fact is made in the final reporting of the **results**. There are good reasons for examining alternate specifications of a model and reporting the **results**;3 however, the primary theme of this paper is that when a specification **search** is undertaken, levels of significance cannot be interpreted in the same way as when a...

...many specifications are openly reported.5 One reason is that adjusting significance levels for the **results** of specification **search** is only recently possible. Some adjustment is possible using the bootstrap procedure described by Veall...

...an alternative method of evaluating significance levels, which is based on viewing the entire specification **search** as the estimator.

This paper describes the problem specification **searches** pose for inference, presents the **results** of some simulations for purposes of illustration, and uses the bootstrapping procedure to give a...

...of statistical significance than a standard t-test. The value of the illustrations of specification **searches** is that they help demonstrate the severity of the problem. The examples presented below illustrate that in most cases, a researcher can undertake a specification **search** and report a statistically significant **result** regardless of whether the variables in a regression equation are actually related. The bootstrap procedure used to



analyze the specification **searches** does provide another way to examine the true statistical significance of empirical **results** . Two different specification **searches** are examined: a "drop insignificant coefficients" **search** and a "biggest t-ratio" **search** . Both are shown to lead to larger than reported standard errors.

#### AN EXAMPLE WITH RANDOM NUMBERS

To illustrate the problem, a specification **search** was undertaken to test the hypothesis that  $Y = f(X_1)$ , using data sets generated by...

...were generated for each variable. One hundred data sets were generated so that the specification **search** procedure could be replicated 100 times. Specification **search** was attempted in two dimensions. First, the intercept and all possible combinations of independent variables...

...two-tailed test was used, since either sign on  $X_1$  is considered an acceptable (publishable) **result** . This specification **search** follows the common practice for empirical researchers to report that variables have been added or...

...their models, due to multicollinearity or other types of problems. The second type of specification **search** was undertaken by looking at only half of the sample; that is, examining the model...

...sample was examined this way, although both halves could have been. This type of specification **search** is done for various reasons with various types of data. In time series, the years...

...dropped from the sample because they behave differently from large firms.

The types of specification **search** done here were quite limited. Frequently, logs and first differences are examined, and the imaginative... using quarterly rather than monthly data. The point is that the two types of specification **search** done here - choosing the best combination of ceteris paribus independent variables, and eliminating some observations - are a small subset of the possible specification **searches** that could be undertaken, but these are two of the commonly used types of specification **search** .

When the regression  $Y = a + bX_1$  was estimated using the one hundred data sets of... $X_6$  were tried along with  $X_1$  in both the small and the large samples. The **result** was that in 17 out of the 100 data sets of random numbers,  $X_1$  was...

...a reported .05 level" in at least one specification. With random numbers one would expect **results** significant at the .05 level in only 5 out of 100 cases, but even a simple specification **search** greatly increases the probability of finding "significant" **results** .<sup>7</sup> In this case our specification **searches** produced "significant" **results** more than three times as often. The implication is that even a simple specification **search** like the one done here can triple the researcher's chance of finding significant **results** , even if no relationship exists, and may give the researcher as much as a 17 percent chance of finding **results** reported as "significant at the 5 percent level." If a more complex **search** is done, the odds obviously improve, and as noted above, one need not just **search** over one data set as was done here.

#### AN EXAMPLE WITH REAL DATA

Gujarati [1988...

...LINE). The full model contains five independent variables which means that if one wants to **search** for the best specification by omitting variables from the full model, there are 25 or...

...eliminates the specification that contains only an intercept, that leaves 31 specifications that can be **searched**. To examine the impact of various specification **searches** on statistical significance, this section focuses on two independent variables, GNP and PRIME. The analysis first assumes that the researcher is **searching** for evidence that GNP has a statistically significant effect on Total Plastic Purchases, and then that the researcher is **searching** for evidence that PRIME has a statistically significant effect.

The first step in the investigation is to estimate the full model which contains all five of the independent variables. The **results** are given in Table 1, and are shown as model number 1. This specification has...

...coefficient of PRIME is not.

This section will consider only the simplest kind of specification **search**: trying out different combinations of independent variables. The remainder of Table 1 shows the **results** from every possible combination of independent variables. In the full model the coefficient on GNP...

...a theoretical paper on the subject, Caudill and Holcombe [1987] show that when a specification **search** is undertaken to look for "statistically significant" **results** associated with a specific variable, the coefficient will be biased away from zero, so it...

...to uncover a specification in which the coefficient of GNP is statistically significant.

A specification **search** on PRIME uncovers even more interesting possibilities. The coefficient estimates on PRIME range from 346...

...81, -0.75, and 0.70 with GNP, STARTS, and UNEMP, respectively. If one is **searching** for a specification to support a particular hypothesis with regard to PRIME, one could pick...

...strong, negative and strong, or anything in between. While one might find a single regression **result** convincing if it showed a coefficient with a t-ratio of 3.75, as the...

...model 3, this particular example shows that if the regression equation was chosen as the **result** of a specification **search**, inferences may be misleading. In this example we can actually see all of the other alternative regression models. In general, the effects of specification **searches** on inferences depend on the type of **search** and the data set **searched**.

(Table Omitted)

Captioned as: Table 1

(Table Omitted)

In the preceding analysis, among other things, we reported the highest and

lowest parameter estimates **resulting** from the specification **searches** . Presenting this range is advocated by Leamer [1983] and is called extreme bounds analysis; it does provide some evidence on the consequences of specification **search** . Veall [1992] extends Leamer's extreme bounds analysis to obtain the entire distribution of the...

...being investigated, and not simply the range, providing more evidence on the consequences of specification **searches** . The next section examines the procedure suggested by Veall.

#### SYSTEMATIC **SEARCH** PROCEDURES

Assume that the researcher begins with an initial model, but has undertaken some specification **search** , presumably due to some dissatisfaction with the **results** of the initial model.<sup>10</sup> Two types of specification **search** rules are examined in this section. The first is the "drop insignificant coefficients" rule.<sup>11</sup> In this **search** , the full model is first estimated and then variables associated with statistically insignificant coefficients (except...

...Then, this reduced model is estimated and the parameter of interest is noted. The second **search** procedure follows the "biggest t-ratio" rule, which involves estimating all regressions which include the...  
...estimate for that variable. Veall [1992] investigates the effects of a third type of specification **search** which involves the stepwise elimination of variables associated with insignificant coefficients.

If one is interested...

...highly collinear relationship. GNP appears to be much more significant after this apparently reasonable specification **search** , and also has a slightly larger estimated effect. If the **search** is undertaken to find the specification with the highest t-ratio on GNP, then model 3 is the **result** , with a t-ratio of 4.15 and a coefficient of 9.69, more than...

...of GNP. Clearly, the estimated magnitude of the GNP effect is changed by the specification **search** . In a theoretical paper on the subject, Caudill and Holcombe [1987] show that specification **search** biases the absolute value of estimated coefficients away from zero, and that effect is clearly visible here.

PRIME presents an interesting case with these same types of specification **searches** . If one uses the .10 level to identify statistically significant coefficients, then the coefficients of...  
...significant, but if the more stringent .05 level is chosen, PRIME becomes significant. Again, this **result** is due to multicollinearity which exists, to some degree, in all data sets. If the...

...away from zero. The example shows that even what appears to be a reasonable specification **search** procedure, like dropping insignificant coefficients, may not produce **results** much different from choosing the specification with the highest t-ratio.

#### AN ASSESSMENT

In an effort to examine the statistical consequences of these common types of specification **searches** in a more rigorous manner, this section employs the bootstrap method of Efron [1982], which...

...of bootstrapping was used in conjunction with model selection and a

different type of specification **search** by Veall [1992]. Once the initial model has been estimated, the residuals,  $ei$ , and the...

...deviation of the 1000 coefficients from all of the data sets produced by bootstrapping. The **result** of this procedure is close to model 5 in Table 1, which was the model selected originally, following the "drop insignificant coefficients" specification **search**. The coefficient is very close, and the t-ratio is slightly lower. When the specification with the highest t-ratio is chosen, the **results** from the 1000 bootstraps produced a mean value of 7.59 with a t-ratio...PRIME is not. This bootstrapping procedure thus can be a way of identifying whether the **result** of a systematic specification **search** is really statistically significant.

Although PRIME is not statistically significant when the mean of 1000...

...statistically significant coefficient on PRIME. The reason the average did not produce a statistically significant **result** when the individual regressions did is that the large range of values on PRIME was distributed almost symmetrically around zero. The coefficients from the specification **searches** in the 1000 data sets ranged in value from -433.14 to 461.97, with...

...35.84. In contrast, the range for the GNP coefficients under the "biggest t-ratio" specification **search** was from -2.88 to 11.04, with a mean of 7.89. Because the...

...statistical significance for GNP.  
(Table Omitted)

Captioned as: TABLE 2

An alternative analysis of bootstrapped **results** is given by Efron and Gong [1983] and Freedman and Navidi [1986]. These authors suggest examining the percentage of bootstrap simulations containing a "significant" **result** for a variable. A different view of the bootstrapped **results** is possible in this framework. Although the bootstrap procedure suggests that GNP is statistically significant but PRIME is not, a specification **search** on this data set using the "biggest t-ratio" criterion will always turn up a specification with a "significant" coefficient on PRIME. Table 2 examines **results** from the regressions run on the 1000 bootstrapped data sets in a **search** for a "significant" coefficient on PRIME, and shows that for every one of the 1000 data **sets**, there is at least one **model** specification in which PRIME is found to be statistically significant at the reported level. Using...

...16 that produced a statistically significant coefficient on PRIME. The next two columns show the **results** of testing for significance at the .05 level. Thus, reading the table, out of the...  
...10 of the specifications have significant coefficients.

One is 10 percent more likely to find **results** reported to be significant at the .10 level even if no relationship exists, and the same type of **results** are shown in the right-most two columns of Table 2. In that case, 100...

...16 specifications with PRIME coefficients reported to be significant at the .10 level. The bootstrapping **results** suggested that PRIME is not truly a statistically significant variable, yet in an examination of...

...from. At the reported .05 level of significance, at least four

specifications reporting "statistically significant" **results** could be found 72.3 percent of the time. Thus, with some specification **search**, a researcher could report many different models supporting the hypothesis that PRIME is a statistically significant determinant of Total Plastic Purchases.

## CONCLUSION

More than a decade ago, Mayer [1980] and Leamer 1983] warned that one cannot take regression **results** and tests of statistical significance at face value. When many specifications of a model have...

...a data set of random numbers that relatively minor and generally accepted techniques of specification **search** can greatly enhance the probability of finding "statistically significant **results**," when no actual relationship is present. Then, using a real-world data set, it demonstrated significant," using common specification **search** techniques and measuring significance with a standard t-test. **Results** like this are reported in academic journals all the time. Leamer [1983] advocated reporting all different plausible specifications to demonstrate the robustness of the **results**, but as the demonstration above has shown, even then it may be easy to report many plausible "significant" **results** for variables that really are not statistically significant.

Unfortunately, there is no simple test like...

...the significance level of a regression coefficient if it is being reported after a specification **search**.<sup>12</sup> However, building on Veall [1992], this paper has further examined a methodology that may...

...to shed some light on the true statistical significance of a coefficient by using the **results** from many bootstrapped data sets. The bootstrap method, introduced in this context by Veall, is examined for two types of specification **search**. A "drop insignificant coefficients" rule and a "biggest t-ratio" are examined. The specification **search** is applied the same way to all of the bootstrapped data sets, producing coefficients on the variable of interest. The **resulting** set of coefficients is then examined to see if its mean is truly statistically significant...

...provide an indication of the true significance of a coefficient, even after a complicated specification **search**. The technique is very general and can be used in to investigate the effects of any specification **search**, but the impact of the specification **search** depends on the type of **search** and the relationships in the data set. In general, standard errors get larger if a specification **search** has taken place, but exactly how much larger must be determined on a case-by...

...of a model is examined. There is a greater than 5 percent probability of finding **results** "significant at the 5 percent level" in one of many specifications, but less than a 5 percent probability of finding significant **results** in all specifications. Beyond that, the type of specification **search** undertaken and the frequency of "significant" **results** will influence the actual significance of the findings.<sup>13</sup> The bootstrapping technique described above can provide some additional insight, but because few researchers are applying Veall's bootstrap method, one must be skeptical of any empirical **results** reported in economics journals. Articles normally do not provide enough information about the entire process that produced the **results** for readers to accurately evaluate their statistical significance.

Footnote:

## NOTES

Footnote:

One of the earliest methods of specification **search** initiated by the power of the computer was stepwise regression. While the technique has fallen out of favor, it is probably more benign than the typical specification **search** leading up to published **results**. Learner [1978] discusses several types of specification **searches**.

For example, in the March 1997 issue of the American Economic Review (the most recent...

...considering the problem of selecting ceteris paribus variables. Holcombe [1989] also considers problems of interpreting **results** after specification **searches**.

6. All empirical work in this paper, including the generation of the random numbers, was...

...be zero). In this situation the only way additional regressors could lead to "statistically significant" **results** for  $X_1$  would be if the variance were reduced sufficiently so that the t-ratios...

...the random samples we generated as evidenced by the increase in the numbers of "significant" **results** when regressors are added and specifications are **searched**.

Footnote:

8. Reporting the highest and lowest estimate from a specification **search** is called extreme bounds analysis and is advocated by Leamer [1983].

9. Some of the...

...type of pretest estimator. Interested readers are directed to Judge et al. [1985, 75]. This **search** is examined here because dropping insignificant coefficients is a common practice in our profession.

12...Reference:

## REFERENCES

Reference:

Caudill, S. B., and Holcombe, R. G. Coefficient Bias due to Specification **Search** in Econometric Models. Atlantic Economic Journal, September 1987, 30-34.

Efron, B. The Jackknife, the...

...Tests and Significance in Econometrics. Journal of Econometrics, 1995, 5-24.

Leamer, E. E. Specification **Searches**: Ad Hoc Inference with Non-Experimental Data. New York: Wiley, 1978.

Reference:

. Let's Take the Con out of Econometrics. American Economic Review, 1983, 31-43. Lovell, M. C. **Data** Mining. Review of Economics and **Statistics** , February 1983, 1-12. Malinvaud, E. Statistical Methods in Econometrics. Chicago: Rand McNally, 1966. Mayer, T. Economics as a Hard Science: Realistic Goal or Wishful Thinking. Economic **Inquiry** , April 1980, 165-78.

Rizzo, M. J. Praxeology and Econometrics: A Critique of Positivist Economics...

37/3,K/31 (Item 21 from file: 15)  
DIALOG(R)File 15:ABI/Inform(R)  
(c) 2005 ProQuest Info&Learning. All rts. reserv.

01715117 03-66107

**The error detection of structural analytical procedures : A simulation study**

Chen, Yining; Leitch, Robert A

Auditing: A Journal of Practice & Theory vl7n2 PP: 36-70 Fall 1998

ISSN: 0278-0380 JRNL CODE: APT

WORD COUNT: 14850

**The error detection of structural analytical procedures : A simulation study**

...TEXT: model which indirectly incorporates information on the structure of an organization's economic activities. The **results** indicate that the performance of each model, with respect to alpha and beta risks, tends...

...have a greater degree of stability in their business and economic activities. In general, our **results** suggest that auditors can improve the **prediction** and error detection capability of **analytical** procedures by using the information inherent in the natural structure of accounting systems which reflect...

...of the audit and are recommended during substantive testing. Many prior studies have emphasized the **significance** of analytical **procedures** in **detecting** financial statement **errors** and improving audit effectiveness (Hylas and Ashton 1982; Biggs and Wild 1984; Tabor and Willis...

...that structural models will generate predicted values significantly closer to simulated values and as a **result** have smaller mean absolute **prediction** errors than several previously examined **analytical** procedures. These procedures include stepwise, ARIMA, X-1, and Martingale expectation models. Structural models directly...

...regressive moving average model founded on the balances of a particular account. The X-11 **model** is a time- **series** **model** which explicitly incorporates trends and seasonality in account balances. The Martingale model simply considers the...Pany (1990), Wilson and Colbert (1989), and Icerman et al.'s (1993) studies which tested **several** error detection **models** without structural relationships and extends Wild's (1987) research which used a structural model for...

...and what information should be included in analytical procedures.

Researchers have examined the performance of **several** statistical **models** such as univariate regression, multivariate regression, univariate time-series, multivariate time- **series** , and X-11 **models** . Generally, research has shown multivariate time- **series** **models** with the greatest information requirements and computational sophistication to be superior in predictive power to...

...1978; Ang et al.1983). In contrast, Lorek et al. (1992) found limitations to these **results** . Knechel (1988), Wilson and Colbert (1989) and Wheeler and Pany (1990) have gone beyond the examination of prediction accuracy and investigated the error detection performance of **statistical analytical** procedures. They found that models requiring more relevant information and greater sophistication produced significantly more...

...an examination of predictive ability and error detection, found that the



use of sophisticated time- **series models** on quarterly data could be questioned from a cost-benefit perspective. In particular, they found...

...structural relationships among accounting numbers (endogenous variables) and relevant exogenous economic variables into econometric time- **series** or prediction **models**. Nonstructural **models**, in contrast, do not explicitly incorporate these relationships. If accounting numbers, along with key exogenous...

...better than multivariate stepwise models. Because he used a sample of only one company, this **conclusion** may be considered tentative. Dzeng (1994) indirectly incorporated structure using a **vector** autoregressive technique with **multiple data series** for a mid-size university and found it to be superior to ARIMA and regression...

...been done to evaluate the structural model's prediction and error detection performance and the **results** are inconclusive. In a related discussion of future audit practice, Elliott (1994, 1995) suggests that...

...the analytical procedures. It has been suggested that the use of external industrial and economic **data** can improve the **predictive** ability of **analytical** procedures (Neter 1980; Lev 1980; Loebbecke and Steinbart 1987; Wild 1987; Allen 1993). These exogenous...

...analytical procedures for three reasons. First, disaggregated data yield a larger sample size, thus increasing **statistical** power. Second, disaggregated **data** are generally influenced less by structural changes in the organization because analysis involving disaggregated data...further investigation is triggered. This positive approach uses  $IEI=0$  as a control point for **statistical analysis**, where E is an error amount. The negative approach, on the other hand, controls the...

...ways to express the existence of an error and the decision risks associated with an **analytical** procedure or a **statistical** testing approach. A summary is provided in table 1. For the positive approach, the risk...

...and beta risks and Type I and Type II errors are defined differently. As a **result**, an effort to reconcile this difference was attempted in SAS No. 39 (AICPA 1981) and...

...of risks.

In summary, the literature suggests there is a need to develop more sophisticated **analytical** procedures that can provide reliable **predictions** of accounting numbers and detect material errors in an audit. Prior research suggests that structural...

...purpose. However, very limited research has been done to evaluate structural model performance, and the **results** have been inconclusive. The following section describes the research method employed. Next, several propositions are...

...a discussion, limitations and final comments. The appendixes provide data acquisition details and sensitivity analysis **results**.

## RESEARCH METHODOLOGY

There are four phases to our study. Our objective for the first phase... alpha and beta risks. We then replicate the above phases for another period of time, **resulting** in the testing of a grand total of 300 simulated companies. Detailed explanations of each...

...sections on research methodology and in appendix A. We further assess the sensitivity of these **results** in appendix B by comparing them to other **results** not fully presented here for a 36-month estimation period.

Phase 1: Data Acquisition and...

...Captioned as: TABLE 1

Phase 2: Simulation Process

The objective of phase 2 is to **simulate** complete **sets** of monthly financial statements for a large number of different companies from the five sets...

...sales. The general models shown below are varied slightly or are simplified based on regression **results** that reflect the characteristics of each of the five organizations and its economic and accounting...

...3 degrees of structure x 5 companies with different sales behavior patterns) different companies are **simulated**. For each company a complete **set** of financial statements is generated. As in previous studies, we assume these statements are correct before errors are seeded. This is then replicated for another time period **resulting** in 300 companies. Phase 3: Expectation Models In phase 3, we use five expectation models...

...improving the predictive ability. Expectation models are based on the first 48 months of simulated **data** and **predictions** are made for the next 12 months. Nonstructural Models (Formula Omitted)

(Formula Omitted)

(Formula Omitted...

...activities. E-R diagrams are used to design many contemporary database accounting systems. As a **result**, these diagrams, along with their associated relational tables, can play a significant role in the...defines the relationships affecting economic resources, such as changes in inventory, explicitly with accounting identities. **Events**, such as cash **collection** and production, are functions of the endogenous and exogenous variables that are likely to affect...

...which the explicit incorporation of the characteristic structure of the organization's transactions into an **analytical** procedure enhances its **predictability** and error detection capability. In summary, Amer (1993) states "from the auditor's perspective, the use of the ER conceptual modeling...will **result** in more effective audits of accounting database processing environments."

(Table Omitted)

Captioned as: TABLE 2...

...A.1. I . Errors are seeded in the 12-month prediction period. 12 Only the **results** for no error (IEI=0) and material errors (IEI=M) are reported in this manuscript...

...more stringent than the use of 0.5 percent of annual sales.<sup>13</sup> As a **result**, our measure of beta risk is more conservative than if we had used 0.5 percent of the annual sales. We analyze our **results** separately so that any effect of using two definitions may be determined. However,

because the...

...in practice, some modify the positive approach to achieve same ex ante and ex post **results** as the negative approach (Smilius 1990, 163-165).<sup>16</sup> PROPOSITIONS AND **RESULTS** We test four propositions related to several factors which are believed to affect the performance...

...ex post analysis of the error detection performance of the analytical procedures and compare our **results** to a nominal large sample ex ante criteria.<sup>16</sup> We assume that central limit theorem...study for two periods of time (1986-1990 and 1989-1993) and we compare the **results** with those generated for a shorter (36-month) estimation period in appendix B. The **results** of both time periods are presented in the tables. The **results** are presented for both the positive and negative testing approaches and for more than one...

...of the models tested and their potential influence on the alpha and beta decision risks **resulting** from the models' utilization. A larger MAPE implies a larger standard deviation of the prediction...

...values (figure 2) for each of the expectation models are also useful for interpreting the **results** for the positive (panel A) and negative (panel B) approaches. The distribution of the predicted...

...lead to a smaller likelihood of an actual amount A being outside the confidence interval **resulting** in a lower level of alpha risk at IEI=0. However, alpha risk also depends...

...likelihood of an actual amount A-E (or A+E) being inside the confidence interval, **resulting** in a greater beta risk ...prediction error (MPE), such as that for the X-11 model, can lead to anomalies **resulting** in very low or high beta risks regardless of the width of the confidence interval ...

...in the mean absolute percentage prediction errors (MAPE) between the structural and other models. The **results** are summarized in table 5.<sup>19</sup> As can be seen, the structural models do not...

...groups across all degrees of structure in panel A of table 5 for 1988-1990. The **results** for 1989-1993 are in the same direction, particularly for sales-driven accounts, but not...  
...stepwise models with respect to their predictive ability for the earlier time period. While the **results** in the later time period are not as significant, they all favor the structural model when compared to the ARIMA, X-11, and Martingale, as shown in table 3. The **results** also can be observed graphically in figure 2 where the actual distributions are shown for...

...model "performed well at predicting the behavior of accounting numbers and was superior to univariate **models** on **several** dimensions (but) the prediction performance of the structural models was not significantly better than that...

...in footnote 8, thus the stepwise equations are not completely contemporaneous. Given Wild's (1987) **results** and this potential indirect incorporation of the structural relationships among accounting and economic variables, the...

...for the positive approach is controlled at IEI=0 to be a =.33. As a **result**, the alpha risk for the positive approach is expected to be the same for all...

...each model. While beta risk is not controlled at  $IEI=M$ , tighter confidence intervals should **result** in less beta risk than for those models with wider confidence intervals for the positive...

...contrast, the negative approach is controlled at  $IEI=M$  to be a  $\alpha=.33$ . As a **result**, the beta risk for the negative approach is expected to be the same for all...

...above, while the alpha risk is not controlled at  $IEI=0$ , tighter confidence intervals should **result** in less alpha risk than for those models with wider confidence intervals for the negative...characteristics of each model as portrayed in figure 2. An auditor can then use these **results** in selecting an analytical procedure (model) and testing approach based on his or her audit...

...structural and stepwise procedures which may make them beneficial for an auditor to use. This **result** holds for both the positive and negative approaches. The comparative combined positive approach risks are...

...following discussion of the alpha and beta risks presented in table 6 is supplemented with **results** of two-tailed t-tests of the propositions of the differences in alpha, beta, and...  
...the authors.

(Table Omitted)

Captioned as: TABLE 6

There are several common patterns in the **results**. For the positive approach, it is apparent from table 6 that the ARIMA and Martingale...

...has a significantly lower beta risk for the positive approach than the structural model. These **results** are consistent with the earlier discussion of alpha and beta risks as seen in figure...

...Martingale models for all degrees of economic stability and for both time periods. Again, these **results** are consistent with our earlier discussion of alpha and beta risks.

From the combined perspective...

...away from the simulated values as illustrated in table 3 and figure 2. As a **result**, the X-11 model frequently concludes that there is an error whether errors are seeded have **resulted** in a small beta risk. It is also apparent that in terms of audit assurance...

...assurance because the risk is controlled at  $M$ , although some in practice adjust positive approach **results** to approximate negative approach **results**. Our negative approach **results** indicate that all five models achieve satisfactory assurance and control the beta risk close to...

...shown here, we test the general proposition that a greater degree of economic structure will **result** in less alpha and beta risks at the noncontrol points and the risks will be...

...to use the logical relationships among confidence intervals, risk and testing approach to explain these **results**. This relationship does in part explain the paradox where less structure can lead to a...

...or a strong seasonal effect. If we assume that a simpler sales behavior pattern would **result** in a smaller confidence interval for both approaches, these **results** would follow from the logical relationships among confidence intervals, risk and testing approach discussed earlier. These **results** also suggest that auditors may choose to select different criteria when using analytical procedures on...  
...which is an indicator of the character of an organization's economic activity.

#### Materiality and **Prediction** Error

As a final **analysis**, we assess the signaling ability of the five analytical procedures. Like Loebbecke and Steinbart (1987...

...follow in part from the use of monthly data as suggested in prior research. These **results** are encouraging, particularly for the structure and stepwise models. DISCUSSION, LIMITATIONS AND CONCLUDING COMMENTS

#### Discussion

Our overall research hypothesis is that structural **analytical** procedures are better **predictors** of account balances and they are better at detecting errors than other commonly suggested procedures. The **results** suggest that the structural model's predictive performance is superior to nonstructural models (X-11...s financial and economic activities as part of the variable selection process. In addition, our **results** clearly suggest that sophisticated analytical procedures, such as the structural and stepwise models, may be...  
...as illustrated in figure 1.

If an auditor's objective is not overall performance, our **results** suggest that the error detection performance of the structural model relative to the other models...

...much dependent on the nature of testing approach and type of decision risk. As a **result**, auditors must select their expectation model in light of the testing approach used and their...

...in most of the positive testing approach cases the ARIMA and Martingale models yield a **significantly** lower alpha risk of **error detection** than the structural **model** (table 6) and the other models (table B.1 in appendix B). In contrast, the stepwise and structural models tend to **result** in lower beta risks compared to the ARIMA and Martingale models (table 6) for the...

...however, this is actually due to the poor prediction performance of the X-11 model **resulting** in prediction errors so large that a material error is so often concluded and there...

...which is concluding that an account is not in material error when it is, our **results** suggest using the stepwise and the structural models in terms of protection against beta...

...confidence intervals.

If an auditor uses the negative testing approach in controlling beta risk, our **results** suggest that all five models tested achieve beta risks close to or lower than the...

...negative approach.

In general, auditors can select the appropriate testing approach and model which will **result** in achieved risks being close to planned levels. From table 6, the negative approach should...the negative approach) can be significantly reduced with increased economic structure.

From another perspective, our **results** support the applicability of a virtual mathematical model of an organization's economic activities which ...

...This, however, is considered appropriate for this study as it is likely the level of **inquiry** of field auditors, e.g., a specific profit center, division, or product line. Second, even...

...needed using actual monthly data for a large number of companies to confirm the simulated **results** of this study. Third, auditors may have some reluctance to use unaudited accounting data. This...

...for example, found significant support for the inclusion of nonfinancial variables in regression and time- **series models** .

3) A generic structural **model** was tested in this study. Practitioners and auditors have access to information about the characteristics...

...able to establish unique structural models for each company that are better than the generic **set** of structural **models** used in this research.

Fifth, we have probably assessed best case scenarios in this paper...

...interrelationships among the accounts and their relationships with exogenous variables. The ARIMA, X-11 time **series** and Martingale **models** may be easier to track with respect to ex ante risks; but we have shown...

...expectation model based on their entire power curve as suggested by Duke et al. (1982).

### **Conclusions and Contributions**

Past research suggests the need for developing sophisticated expectation models like structural models the **results** have been inconclusive. In this research we build and test a generic structural **model** based on a general **set** of entities ( **events** , resources and agents) and their endogenous and exogenous relationships. It incorporates the logical structure of...

...requires greater computational skill than ARIMA, X- 11 and Martingale models. We test all the **models** on complete **sets** of **simulated** monthly financial statements. Moreover, we test the models using both the positive and negative approaches...

...lower combined alpha and beta risks than the other models for both testing approaches. The **results** for the positive approach suggest that if an auditor wants to minimize the risk of...

...economic events and relevant endogenous variables, such as the structural or the stepwise model. The **results** for the negative approach also favor the structural or the stepwise models. With the lower...  
...within that set which minimizes his or her alpha risk and extra audit effort, our **results** clearly favor the use of either the stepwise or the structural analytical procedure coupled with the negative testing approach.

From the practicing auditor's perspective, the **results** here show that a generic structural model is no better than a stepwise model. This...

...is reasonable assurance that the account is not in material error. From another perspective, our **results** clearly show that practicing auditors who rely on the balance from the previous period, as expressed by the Martingale model as a basis for **analytical** procedures, can greatly improve their **prediction** and error detection capabilities by using the natural structure of the accounting system which reflects the organization's business and economic activities.

Finally, the **results** presented here are based on a generic set of structural equations. In practice, auditors have...

...In contrast, Dzeng (1994) used a monthly data set and its variance-covariance matrix to **simulate** 100 **sets** of data for four time series.

Footnote:

The five companies are selected from the list...

...cs is to add uniform control to the simulation process so we can assess the **results**. All the stochastic equations used in the simulation model contain, not only an  $e$  for...random variables in each equation is much more complex in reality than that which would **result** from a simple normally distributed random variable at the end of each equation. Therefore, each stochastically **simulated** number is a function of **several** random variables including those exogenous to the organization.

Footnote:

The use of 0.5 and...

...small errors using disaggregated data. An error is seeded in each month so that the **results** are less sensitive to random occurrences for each month and the percentages represent a larger...

...procedures.

Footnote:

15 Smieliauskas (1990, 163-165) shows in an appendix how the ex post **results** of the positive approach are modified to achieve the ex post **results** of the negative approach. We do not present the modified positive approach here; we simply recognize that those **results** would be the same as the negative **results** ex ante and ex post.

16 Differences between ex ante and ex post analysis are...

...are 1.9656 and 1.9356 for the structural and stepwise models, respectively. As a **result**, we conclude that there is no apparent autocorrelation of the residuals and that both models...

...Interest Expense" in the original data for Betz Laboratories, Inc. and Cooper Tire & Rubber. This **resulted** in large standard deviations of the residuals  $e$  for the simulation models based on these two companies. This may have **resulted** in small denominators which were used to calculate MAPE. As a **result**, large prediction errors occurred for this time period for nonsales driven accounts. The multiples of...final panel showing the randomized curve-fitted monthly data in figure A.2. As a **result**, the natural structure of the simulated companies may be slightly less than

observed in practice feasibility of using **statistical** -based models as **analytical** procedures. Unpublished monograph, Florida State University.

Johnson, J. R., R. A. Leitch, and J. Neter...

...of Practice & Theory 7 (Spring): 59-73.

Reference:

Knechel, W. R. 1988. The effectiveness of **statistical analytical** review as a substantive auditing procedure: A simulation analysis. The Accounting Review 63 (January): 74...

...K. S., B. C. Branson, and R. C. Icerman. 1992. On the use of time-**series models** as analytical procedures. Auditing: A Journal of Practice & Theory 11 (Fall): 66-87. Neter, J...

...graphically depicts the time-series behavior of each company's sales for the 20 quarters **ending** December 31, 1990. Second, for each of the five companies we obtain quarterly data from...



37/3,K/36 (Item 26 from file: 15)  
DIALOG(R)File 15:ABI/Inform(R)  
(c) 2005 ProQuest Info&Learning. All rts. reserv.

01472449 01-23437

Discovery and representation of casual relationships in MIS research: A  
methodological framework

Lee, Byungtae; Barua, Anitesh; Whinston, Andrew B  
MIS Quarterly v21n1 PP: 109-136 Mar 1997  
ISSN: 0276-7783 JRNL CODE: MIS  
WORD COUNT: 14661

Discovery and representation of casual relationships in MIS research: A  
methodological framework

...ABSTRACT: MIS models. Since richer models require more flexible tools and techniques, the representational shortcomings and **statistical** pitfalls of factor- **analytic** methods commonly deployed in empirical research are described. It is suggested that weak exploratory phase...

...TEXT: richer models require more flexible tools and techniques, the paper describes the representational shortcomings and **statistical** pitfalls of factor- **analytic** methods commonly deployed in empirical research. We suggest that weak exploratory phase tools and approaches...

...leads to success, a simplistic model of reengineering payoff through radical change can provide misleading **results**. Incorporating "why" or "when" aspects of a relationship in a model **results** in an enriched and more accurate representation of the problem of interest. Along these lines ...

...in accordance with his/her theory. We find that existing tools such as exploratory factor **analysis** are both conceptually and **statistically** inadequate for this purpose and that incorrect modeling assumptions and measurement problems can carry over...Sobel (1995)). The ontological aspect of causality involves the characterization of the relationship between a **cause** and its **effect** (s). Three elements in the relation of **cause** and **effect** have been identified: temporal precedence, contiguity in time and space, and constant conjunction (Hume 1977...

...thought are called "realists" or "natural necessity theorists." They focus on mechanisms or processes connecting **cause** and **effect** (i.e., the "why" and "when" aspects), while regularity theorists emphasize scientific laws or universal...

...possibility of contemporaneous causation and events that may take place between the occurrences of a **cause** and its **effect**. In fact, Simon (1953) explicitly and purposely avoids the temporal aspect of Humean causality in of the relationship between a **cause** and its **effect** and is conceptually different from Hume's temporal priority. Sobel suggests that a "manipulative account...

...the assumption that there is a relationship between output quantity produced by a firm and **input** expenditures including IT investments. A change of **input** levels **results** in a change of output quantities, and the assumption is that the **inputs** are exogenous. By contrast, the Barua and Lee (1997) IT productivity model assumes that firms change the levels of **inputs** depending on the **input** prices. Thus, as the price of computing falls, firms use more IT. The Barua and Lee model is a system of  $n + 1$  relationships, one for the **inputoutput** linkage, and  $n$  relationships describing how the firms choose  $n$  **inputs** depending on **input** prices and

other factors.

A similar position is taken by Sobel (1995) in sociology, who...research studies, 69% (62/90) proposed causal models (i.e., developed testable hypothesis linking a **cause** and its **effect**). "Survey" and "instrument development" research is classified under "field studies" but under non-causal modeling...through additional relationships will help explain how agents choose their actions, which in turn affect **outcomes** of interest.

Isolation between two modeling worlds

Two separate worlds of causal modeling have been...

...be large at the exploratory stage, this two-step analysis can lead to significantly incorrect **results**. Although LISREL allows the estimation of coefficients of the structural equations without the explicit calculation ...factors. Since these assumptions are not empirically tested, their violations and consequential distortions of the **results** go unnoticed. One study relaxed some restrictions on error terms (Joreskog and Sorbom 1982), but...

...the above model. In fact, a more convincing scenario may depict socioeconomic status as an **outcome** of education, income, and occupation (Glymour et al. 1987). Then the causal dependencies are the...  
...with very different implications for estimation. Since causality implies an inherently asymmetric relationship between a **cause** and its **effect**, reversing the direction of causality **results** in a different **model** with the same **set** of variables. Similar analogies can be applied to the MIS domain. For example, MIS personnel...that is outside the scope of LISREL.

Linking the two research phases with LISREL: the **search** for richer models

The claim has been made that an important contribution of experimental studies...

...a developing field than in a well-established discipline. Consider a commonly encountered relationship between **input** variables such as IT investment, labor, etc., and firm performance measures such as market share ...providing better service to customers.

LISREL takes advantage of model fitting techniques in exploring alternative **models**. For a given **set** of data, LISREL maximizes the fit between data and the model. If the model does...

...to new theory development. An exploratory tool should contain provisions to account for such possibilities.

**Searching** for a New Approach to Exploratory Research

While we are in total agreement with the...

...TETRAD represents a proposed causal model as a labeled directed graph. Such a graph uniquely **determines** the structural equations of the statistical **model**, while undirected paths represent **correlation** without causality (e.g., correlation among **error** terms). For example, consider three normalized variables X, Y and Z. Suppose a hypothesized causal...

...two models have different implications for TETRAD.

Advantages of TETRAD

Non-parametric analysis: Unlike factor **analytic** approaches, there is no **statistical** parameter estimation for a hypothesized causal model in TETRAD. Thus, TETRAD is free of many...derived under the assumption of multivariate normality of observed variables and may lead to misleading **results** when the assumption is violated. Second, outliers can change the magnitude of TETRAD differences and lead to incorrect model selection. Another problem recognized by the developers of TETRAD involves **multiple** testing on each **model** (Glymour et al. 1987). Bollen (1990) has recommended remedies for each of these problems. For...

...also notes that TETRAD does not contain provisions for specifying correlations between exogenous variables in **multiple** regression **models** and hence cannot help in model modification for exogenous variables. However, by their very definition...

...test (Hausman 1978) accomplishes this objective and has been used to show the endogeneity of **input** variables such as IT investment by (Barua and Lee 1997).

Examples highlighting TETRAD's advantages...

...problem provided the exploratory stage research shows that such links do not exist). Using a **series** of **simulation** studies with sample sizes varying from 200 to 2,000, it has been demonstrated that...where the factor analytic representation requirements force the researcher to reverse the direction of causality.

**Detecting** anchoring and adjustment: A **correlated errors model** was used for a factor "Authoritarianism-Conservatism" with five indicators measured on a five point...

...of interest. We suggest adding several new elements, including the conceptual assessment of exogeneity of **predictors** in the **theoretical** model, a manipulative account of proposed relationships, as well as a consideration of enriching the...

...model itself. Having formulated a baseline model, the researcher must ask the question: Can we **theoretically** justify the **predictors** in the current model as exogenous? A number of examples have been provided where variables...

...Step 2: This involves the representation of a preliminary causal model. Following step 1 may **result** in a model that is outside the domain of factor analytic approaches (e.g., a...require that such linkages be theoretically justified.

The second situation involves unintentional causality between indicators **resulting** from anchoring and adjustment on the part of the respondents. This possibility should not be...

...increase in the scope of the model), then we return to step 1. There are **several** possible **scenarios** here. First, the researcher may have to enhance the theory by identifying more agent, team...

...than adding extra variables to an initial model. More often than not, this step will **result** in additional relationships arising from deeper insights into the processes connecting the variables of interest...

...that are not reliable and return to step 3 for new data collection.

If the **outcome** of applying TETRAD suggests that no additional variables

be added (i.e., either the initial...at this stage depend on the model at hand. For example, the proposed procedure may **result** in a model out of LISREL's scope. For instance, there may be direct causal...

...assuming the variable to be exogenous and proceeding with a simplified analysis of the model.

### **Conclusion**

Being a relatively young discipline is a mixed blessing for MIS. On the one hand...

...the analogy of exploration for natural resources, we suggested that a richer theory and its **resulting** model enable the researcher to focus only on specific geographical areas with a high likelihood...not isolated recommendations; rather, a manipulative account involving how factors such as incentives affect behavior **results** in a clear understanding of exogenous and endogenous variables in a model.

In representing causal...Sloan School of Management, Cambridge, MA, 1990.

Costner, H. and Schoenberg, R. "Diagnosing Ills in **Multiple** Indicator **Models** ," in Structural Equation **Models** in the Social Sciences, A. Goldberger and O. Duncan (eds.), Seminar Press, New York, 1973...

...M. A., Price, R. L., and Marrs, D. W. "Linking Theory Construction and Theory Testing: **Models** with **Multiple** Indicators of Latent Variables," Academy of Management Review (11:1), 1986, pp. 128-144.

Hume...

Set	Items	Description
S1	7360007	IDENTIF? OR LOCAT? OR NARROW?()DOWN? OR FIND? OR RETRIEV? - OR TRACK? OR REVEAL? OR ISOLAT?
S2	2329080	ASCERTAIN? OR DISCERN? OR SIFT? OR FILTER? OR RECOGN? OR D- ISTINGUISH? OR UNCOVER?
S3	2506877	SINGL?()OUT OR PINPOINT? OR SORT??? OR INDICAT? OR DESIGNA- T? OR TURN?()UP OR UNMASK? OR BACKPROPAGAT?
S4	2884435	DETECT? OR DISCOVER? OR UNEARTH? OR EXPOSE? OR EXPOSING? OR CULL? OR FERRET? OR DETERMIN?
S5	1268068	S1:S4(7N) (METHOD? OR SYSTEM? OR PROCESS?? OR PROCEDUR? OR - TECHNIQUE? OR MODE? ?)
S6	44878	S5(7N) (PATTERN? OR SIMILAR? OR LIKENESS? OR RELATIONSHIP? - OR CONNECTION? OR CONCATENAT? OR EXPECT?)
S7	37932	S5(7N) (KINSHIP? OR LINK? OR AFFILIAT? OR CORRESPOND? OR MA- TCH? OR SIGNIFICAN? OR LIKELIHOOD?)
S8	15922	S5(7N) (ASSOCIATION? OR CORRELAT? OR COINCIDEN? OR TREND? OR TRAJECTOR? OR COINCID? OR ANTICIPAT?)
S9	2060	S6:S8(7N) (INFLUEN? OR IMPACT? OR SWAY? OR LEVERAG? OR BIAS? OR SLANT? OR PROPENSIT?)
S10	12	S6:S8(7N) (AFFECTATION? OR IMPING? OR CROSS()POLLEN? OR TAI- NT? OR SCREW?()UP OR SNAFU?)
S11	1333	S6:S8(7N) (ERROR? OR FAULT? OR GLITCH? OR ADULTERAT? OR WAR- P? OR SPOIL? OR BLEEDTHRU? OR BLEEDTHROUGH? OR BLEED?() (THRU - OR THROUGH?)
S12	248	S6:S8(7N) (SKEW? OR CONFOUND? OR MONKEY()WRENCH? OR PERTURB? OR MANIPULAT? OR CORRUPT?)
S13	372	S6:S8(7N) (UNFORESEEN? OR UNKNOWN? OR UNPREDICTAB? OR UNACC- OUNT? OR UNEXPECT? OR UNANTICIPAT? OR OVERLOOK? OR INADVERTEN- T?)
S14	3949	S9:S13
S15	2862	S14 AND (OUTCOME? OR RESULT? OR CONCLUSION? OR ENDRESULT? - OR EVENTUALIT? OR CAUSE?(2W)EFFECT?)
S16	613	S14 AND (DESTINAT? OR FINALE? OR COMPLETION? OR CESSATION? OR ENDING? OR PROXIMAT?()CAUSE?)
S17	165	S14 AND (TERMINATION? OR AFTERMATH? OR UPSHOT? OR FALLOUT?)
S18	167	S14 AND (REPERCUSSION? OR CULMINAT? OR RAMIFICAT? OR HISTO- GRA?)
S19	2927	S15:S18
S20	503	S19 AND (SET OR SETS OR GROUP? OR CLUSTER? OR ARRAY? OR AS- SEMBL? OR PLURAL? OR SEVERAL? OR MULTIP? OR MULTIT? OR COLLEC- TION? OR ASSORTMENT? OR NUMEROUS? OR SERIES? OR NODE?) (5N) (CO- NDITION? OR HYPOTHE? OR THEORET? OR PRETEND? OR SIMULAT?)
S21	298	S19 AND (SET OR SETS OR GROUP? OR CLUSTER? OR ARRAY? OR AS- SEMBL? OR PLURAL? OR SEVERAL? OR MULTIP? OR MULTIT? OR COLLEC- TION? OR ASSORTMENT? OR NUMEROUS? OR SERIES? OR NODE?) (5N) (EN- VIRONMENT? OR SITUATION? OR CONTINGEN? OR QUALIFICATION?)
S22	175	S19 AND (SET OR SETS OR GROUP? OR CLUSTER? OR ARRAY? OR AS- SEMBL? OR PLURAL? OR SEVERAL? OR MULTIP? OR MULTIT? OR COLLEC- TION? OR ASSORTMENT? OR NUMEROUS? OR SERIES? OR NODE?) (5N) (ST- IPULATION? OR EVENT? OR CIRCUMSTAN? OR HAPPENING?)
S23	151	S19 AND (SET OR SETS OR GROUP? OR CLUSTER? OR ARRAY? OR AS- SEMBL? OR PLURAL? OR SEVERAL? OR MULTIP? OR MULTIT? OR COLLEC- TION? OR ASSORTMENT? OR NUMEROUS? OR SERIES? OR NODE?) (5N) (PH- ENOMEN? OR INCIDENT? OR OCCURRENC? OR EPISODE? OR SCENARIO? OR VECTOR?)
S24	636	S19 AND (SET OR SETS OR GROUP? OR CLUSTER? OR ARRAY? OR AS- SEMBL? OR PLURAL? OR SEVERAL? OR MULTIP? OR MULTIT? OR COLLEC- TION? OR ASSORTMENT? OR NUMEROUS? OR SERIES? OR NODE?) (5N) (MO- DEL? ? OR (TEST OR INSTANT OR REFERENCE OR IDEAL?) () (CASE? OR SET OR SETS)
S25	1122	S20:S24

S26 823 S25 AND (QUERY? OR QUERIE? OR REQUEST? OR INTERROG? OR INQ-  
 UIR? OR SEARCH? OR RETRIEV? OR INPUT? OR INTERFAC?)  
 S27 557 S25 AND (STATIST? OR PROBABILIT? OR PREDICT?) (5N) (ANALY? OR  
 DATA? OR THEOR? OR FORMULA? OR ALGORITHM?)  
 S28 419 S26 AND S27  
 S29 53 S28 AND S1:S27/TI  
 S30 45 S29 AND PY<2002  
 S31 42 RD (unique items)  
 File 275:Gale Group Computer DB(TM) 1983-2005/Aug 15  
 (c) 2005 The Gale Group  
 File 369:New Scientist 1994-2005/May W5  
 (c) 2005 Reed Business Information Ltd.  
 File 370:Science 1996-1999/Jul W3  
 (c) 1999 AAAS  
 File 484:Periodical Abs Plustext 1986-2005/Aug W1  
 (c) 2005 ProQuest  
 File 553:Wilson Bus. Abs. FullText 1982-2004/Dec  
 (c) 2005 The HW Wilson Co  
 File 613:PR Newswire 1999-2005/Aug 15  
 (c) 2005 PR Newswire Association Inc  
 File 621:Gale Group New Prod.Annou.(R) 1985-2005/Aug 15  
 (c) 2005 The Gale Group  
 File 624:McGraw-Hill Publications 1985-2005/Aug 15  
 (c) 2005 McGraw-Hill Co. Inc  
 File 634:San Jose Mercury Jun 1985-2005/Aug 13  
 (c) 2005 San Jose Mercury News  
 File 635:Business Dateline(R) 1985-2005/Aug 15  
 (c) 2005 ProQuest Info&Learning  
 File 636:Gale Group Newsletter DB(TM) 1987-2005/Aug 12  
 (c) 2005 The Gale Group  
 File 647:CMP Computer Fulltext 1988-2005/Jul W5  
 (c) 2005 CMP Media, LLC  
 File 674:Computer News Fulltext 1989-2005/Aug W1  
 (c) 2005 IDG Communications  
 File 696:DIALOG Telecom. Newsletters 1995-2005/Aug 12  
 (c) 2005 Dialog  
 File 810:Business Wire 1986-1999/Feb 28  
 (c) 1999 Business Wire  
 File 813:PR Newswire 1987-1999/Apr 30  
 (c) 1999 PR Newswire Association Inc

31/3,K/2 (Item 2 from file: 275)  
DIALOG(R)File 275:Gale Group Computer DB(TM)  
(c) 2005 The Gale Group. All rts. reserv.

01791896 SUPPLIER NUMBER: 16943999 (USE FORMAT 7 OR 9 FOR FULL TEXT)  
**Box-Jenkins forecast model identification . (includes a related article  
the mathematics behind the back-propagation technique described) (Cover  
Story) (Technical)**  
Reynolds, Steven B.; Mellichamp, Joseph M.; Smith, Robert E.  
AI Expert, v10, n6, p15(14)  
June, 1995  
DOCUMENT TYPE: Cover Story Technical ISSN: 0888-3785 LANGUAGE:  
ENGLISH RECORD TYPE: FULLTEXT; ABSTRACT  
WORD COUNT: 9885 LINE COUNT: 00846

**Box-Jenkins forecast model identification . (includes a related article  
the mathematics behind the back-propagation technique described) (Cover  
Story) (Technical)**

ABSTRACT: Box-Jenkins ARIMA forecasting models produce better forecasts than other time- **series** forecasting methods, but the **model** identification task is complex. The Box-Jenkins **model** identification problem for nonseasonal time **series** is addressed here using three-layer back-propagation artificial neural networks. Two neural networks are...  
...differences required to render a time series stationary, and the second identifies appropriate Box-Jenkins **models** for stationary **series** indicated by the first network. Neural networks are shown to be very effective in identifying appropriate Box-Jenkins ARIMA **models** for nonseasonal time **series** ; **models** recommended were consistently better in terms of quality and fit than those indicated by a...

In this research, the Box-Jenkins **model** identification problem for nonseasonal time **series** was addressed using three-layer back-propagation artificial neural networks. We constructed two neural networks...

...a time series stationary. We designed a second neural network to identify appropriate Box-Jenkins **models** for stationary **series** indicated by the first network.

The **results** of the research indicate that neural networks are very effective in identifying appropriate Box-Jenkins ARIMA **models** nonseasonal time **series** . Tests of the networks' performance in identifying appropriate **models** for **series** of various lengths confirmed that the models recommended were consistently better in terms of quality...

...a time series from its own past values. One of the most consistently successful time- **series** forecasting **models** was developed by George E. P. Box and Gwilym M. Jenkins.

Box-Jenkins ARIMA forecasting...

...the Box-Jenkins approach to time-series forecasting yields forecasts that are superior to those **resulting** from other standard time-series forecasting procedures. As a **result** , the method has received much attention and has been examined frequently in the literature. However...  
...the difficulties associated with model identification.

Vandaele states, "Identification is the key to time-series **model** building." Box-Jenkins nonseasonal time- **series** **models** consist of three components: an autoregressive (AR) component, a moving average (MA) component, and an...

...that it is sometimes advantageous to include both AR and MA terms in a

time- **series model** . Box-Jenkins **models** are typically denoted ARIMA (p,d,q) where p indicates the order of the autoregression...judiciously consider the variance of the time series ACF and PACF while conducting Box-Jenkins **model identification** .

Another factor that clouds the **model identification process** is human **bias** in **pattern recognition** . Lopes **indicates** that when presented with multiple plots of data, people frequently fail to detect existent patterns...

...the Box-Jenkins model identification task is not surprising.

The autoregressive and moving average time- **series models** described earlier are both predicated on the assumption that the underlying time series is stationary...

...the most appropriate for a given series arises since the estimated AGF and PACF patterns **resulting** from actual **series** never exactly match **theoretical** patterns. Box and Jenkins explain why perfect matches should not be expected: "... moderately large estimated...linearly separable. Fortunately, Kolmogorov's theorem states that as long as the elements of an **input** vector lie between zero and one, a three-layer network can perfectly map any real...

...complex decision regions." Rumelhart, Hinton, and Williams also confirm that multilayered networks can always relate **inputs** to appropriate outputs.

In addition to excellent mapping abilities, neural networks have good generalization characteristics...

...problems ... and may complement statistical and expert systems approaches to involved pattern recognition." She lists **several conditions** where neural networks may ...examples of decisions are available and in problems where a large number of attributes describe **inputs** . The author also suggests neural networks may be more advantageous in problems where uncertainty exists...

...that neural networks are better suited in cases where a good deal of noise confounds **statistical analysis** . Additionally, she contends that neural networks may be preferable in problems where outliers exist and...

...was to apply neural network technology to the task of identifying appropriate Box-Jenkins forecasting **models** for time **series** . Two neural networks were developed to accomplish this task. One neural network (Stage 1) was...

...a system of coupled networks is often an efficient approach to solving complex problems.

#### NONSTATIONARY **SERIES**

The Box-Jenkins ARIMA **model** identification method assumes the statistical process underlying a time series of interest is stationary. That...

...nonstationary ACFS were simulated on an IBM mainframe computer using a FORTRAN program. Forty-eight **series** of length of 50 were **simulated** . A basic assumption of the Box-Jenkins forecasting method is that the random shocks in...

...because it is generally agreed that 50 is the minimum required sample size.

Since nonstationarity **results** only from series with autoregressive parameters, Ar(1), AR(2), and ARMA(1,1) nonstationary **series** were



**simulated** . The 48 **series** were **input** to the Execustat statistical software package, and the ACFs and standard errors were computed for...

...to add robustness to the network training process, the 48 ACF/standard error vectors were **input** to a BASIC program for noise addition. Caudill describes the benefits of adding noise to network **inputs** : "Applying a little noise to the training set has a highly desirable effect of producing a network more robust to noisy **inputs** after training .... The exact type and amount of noise depends on your data, expected level...

...ACF spike was a uniformly distributed random value between -0.03 and 0.03. ACFs **resulting** from nonstationary series often have very high autocorrelations. In fact, the initial ACF spikes commonly...

...be trained to distinguish between the two series types. To obtain the required sample, stationary **series** were **simulated** . As previously discussed, six basic Box-Jenkins models exist. By varying the coefficients of the model parameters, 19 tifiable subcases **result** . Table 3 details the 19 subcases. Ten examples of each of the 19 subcases were...

...set.

TABULAR DATA 3 OMITTED|

The combination of the nonstationary and stationary ACF/standard error **vectors** comprised a training **set** of 382 **vectors** . Each of the 12-element (six ACF and six standard error) training vectors was assigned ...

...targets of 10 while the stationary vectors were assigned a target of 01. The 382 **vector set** was **input** into the PlaNet **simulator** . Backpropagation networks with one, three, and six hidden nodes were trained. All three networks converged Caudill. Each test vector was presented to the weight matrices **resulting** from the network training phase, and the outputs were computed. The output element with the...

...was the one that fired with the greatest strength, thus indicating which of the two **series** types the **input vector** most closely resembled. The **results** of the test indicated that the networks with three and six hidden nodes correctly mapped...

...appropriate output category. The network with one hidden node correctly mapped 59 of the 60 **test cases** .

The 60 test **series** were then **input** to the Forecast Pro expert system for evaluation. The Forecast Pro expert system incorrectly recommended undifferenced models for three of the 30 nonstationary **series** and differenced **models** for four of the 30 stationary series. Thus, where the neural network obtained perfect **results** , the expert system either underdifferenced or overdifferenced seven (12%) of the test cases.

Further analysis...

...Hinton diagrams was then conducted to determine if an even smaller network could produce similar **results** . Based on this analysis, new networks were constructed with only the six ACF elements as **inputs** . Mozer and Smolensky termed this sort of network pruning "skeletonization." They contend that by eliminating **input** and hidden units that serve no purpose, networks can be trained more quickly and improved generalization often **results** . Work by Apolloni, Avonzini, Cesa-Bianchi, and Ronchini confirmed that trimming network connections with low weights produced improved **results** .

RELATED ARTICLE: Mathematics

The power of the backpropagation network is based on the adjustment of

...layer. Application of the GDR begins with the processing of signals through the network. The **inputs** to the network are denoted as  $X_{pi}$  where  $p$  denotes the  $p$ th **input** training vector and  $i$  indicates  $i$ th **input** node. Typically, network **inputs** consist of either raw or transformed data that serve as the basis for the eventual...

...node. The bias node has a constant signal value of one and, like the other **inputs**, has an associated adjustable weight. Freeman and Skapura state that experience shows that the inclusion of a bias node often aids in achieving acceptable network error convergence.

**Input** signals are transmitted forward through the network to the hidden layer of nodes. At the hidden layer, the weighted **input** signal values are summed across all **inputs**. This value is denoted as:

Mathematical Expression Omitted where Mathematical Expression Omitted is the weight associated with the connection between **input**  $i$  and hidden node  $j$ ,  $X_{pi}$  is defined as before and Mathematical Expression ...

...sigmoidal (or logistic) function:

Mathematical Expression Omitted was found to be quite useful. The value **resulting** from the application of the sigmoid activation function at the hidden layer constitutes the signal...

...is the sigmoid function.

The same procedure is followed at the output layer. The weighted **inputs** from the proceeding (hidden) layer are evaluated in a sigmoid function and the output signal is computed. These output values are denoted as Mathematical Expression Omitted.

After the **inputs** are propagated forward through the network, the weight adjustment process begins. For each training vector...

...GDR algorithm to take larger steps down the error surface. Thus, a large  $\eta$  may **result** in faster convergence. However, if  $\eta$  is too large, the error may not converge to...is the incorporation of a momentum factor in the Generalized Delta Rule. The momentum factor **results** in the addition of a fraction of the previous weight change to the next weight...

...Networks with one, three, and six hidden units were trained and tested with the smaller **input** vectors. Remarkably, the **input** network with a single hidden **node** perfectly categorized the 60 test vectors.

#### STATIONARY SERIES MODEL ID

This section describes the more complex of the two networks discussed earlier--the development...

...development of the data needed to train the neural network. The approach taken was to **simulate multiple series** of length 50 for each of the six basic Box-Jenkins processes (random,  $Ar(1)$ ...  
...computer.

Thirty training patterns were simulated for each of the 19 subcases discussed earlier. The **resulting** 570 **simulated series** were **input** to Execustat to compute the ACFS, the PACFs, and the associated standard errors for each...

...target for each training vector indicated which ARIMA process was used to generate the underlying **series** upon which the training **vectors** were based. A dense mapping approach was used so that each of the six basic...

...network configurations with 12, 18, 36, and 48 hidden nodes were trained

with the 570 **input** vectors using the Generalized Delta Rule and the Planet simulation package. All four network schemes...

...pattern set.

The trained networks were tested with 20 additional simulated ACF/standard error/ PACF **vectors** . The **set** of 20 included two random series and one series of each of the other 18...

...output element that fired with the greatest strength indicated which of the six patterns the **input** vector most closely resembled.

The output of the tests indicated problems. The networks yielded only ...

...problem was discovered involving the simulated test vectors. A thorough analysis of the 20 test **vectors** indicated that **several** of the test series were either close calls or actually better modeled by processes other...

...them. These findings indicated that an alternative approach was needed in training the network. Since **simulated series** were sometimes better fit by **models** other than those used to simulate them, it was concluded that a training file based...

...analyses were conducted to examine the extent of this apparent problem. An additional 57 time **series** were **simulated** . This **set** of data included three series from each of the 19 subcases described earlier. Each of the 57 series was **input** to Execustat and fit to all six of the basic Box-Jenkins models. The analysis...

...says that training data is seldom without errors and one must be careful about the **ramifications** of simply throwing out the data. Even flawed data often contains important information that is...

...effort to overcome the difficulties of incorrect training data was to significantly expand the training **set** . Additional series were **simulated** to enlarge the training **set** to 1,270 **vectors** . The training and testing of networks with 18, 36, 72, 90, and 108 hidden nodes...

...examined via a Hinton diagram. This exercise indicated that the standard error portion of the **input** vectors made little contribution to network performance. Lawrence confirms that neural networks learn to ignore **inputs** that have little or nothing to do with the problem.

In addition to the standard error values, the number of ACF and PACF **inputs** used in the **input** training vectors was considered. The first few spikes of the ACF and PACF plots are...

...contain statistical noise that can cloud the model identification task. Based on these facts, the **input** training vectors were altered accordingly. The new **inputs** consisted of the first six ACF elements, the first six PACF elements, and the six target elements described earlier.

This reduction in the **input** vector size still did not address, however, the problem of erroneous mappings in the training...

...Pouliot, Bengio, and Agin, and Caudill indicated that an effective way of expanding a training **set** is to generate training **vectors** by adding stochastically generated noise to **input** elements of known correct training examples. Lawrence states that a manufactured training set that contains basic known patterns can provide substantially improved **results** .

Since experience clearly indicated that simulated vectors sometimes mapped to processes other than those used to generate them, additional training **inputs** for basic known patterns were manufactured. The

additional training vectors were manufactured by creating sanitized versions of the unique general patterns that were evident in the **set** of 1,270 **simulated vectors** described earlier. This sanitation of the patterns was based on the theoretical patterns illustrated by...

...All the networks successfully converged to less than 4% error within 50 training cycles. Each **resulting** trained network was tested with 57 simulated vectors (three vectors representing each of the 19...

...model, these criteria were applied to identify other models that were "good" fits to the **series**. A good fitting **model** was defined as one that met the above criteria and had an RMSE within 10...

...the best fitting model. Based on these criteria, the neural network that produced the best **results** was the network with 18 hidden nodes.

The same 57 time series were also evaluated...

...Forecast Pro expert system. Tables 4 and 5 summarize the performance of the model recommendations **resulting** from the two systems. Table 4 summarizes the performance of the Stage 2 neural network in identifying appropriate stationary Box-Jenkins **models** for the 57 time **series**. Note the network identified the best fitting **models** for 32 of the time **series**. In addition, "good" **models** were indicated for 12 **series**. When eight of the **models** recommended by the network were fit to the data, insignificant coefficients (p-value greater than .10) **resulted**. Note, however, that four of the eight models had reasonably small p-values of less...

...fitting" models.

TABULAR DATA 4 & 5 OMITTED|

Table 5 summarizes the Box-Jenkins model recommendations **resulting** from the **input** of the 57 time series to the Forecast Pro expert system. As shown, the Forecast...

...the expert system also recommended nonparsimonious models (more than two parameters) for three of the **series**. The remaining seven **model** recommendations indicated models with simple differencing. The differenced models not only produced inferior RMSEs, but...

...5 clearly show that the neural network more often identified the best fitting basic ARIMA **model** for the 57 **series** evaluated. Further, as shown, the neural network was marginally better than the Forecast Pro expert models identified

#### RESULTS

Testing of the neural networks with actual time series was conducted in two phases. In the first phase, the networks were presented with ACF and PACF **inputs resulting** from actual time series that were of approximately the same length as those series used...

...was conducted to assess the generalization abilities of the networks in identifying appropriate Box-Jenkins **models** for actual time **series** of greater lengths.

#### SIMILAR LENGTH TIME SERIES

Monthly sales data were collected from a large...

...of regular differences required to achieve stationarity. The first step in this process was the **input** of the 36 time series to Execustat for computation of the ACF values. These **results** were arranged in vectors containing the first six ACF elements and a six-position dummy target for presentation to the Stage 1 neural network. The dummy target used for all **inputs** was 10. The dummy target was included strictly to accommodate the Planet simulation package format...

...was applied in a building block approach similar to the one described by Hoff. Each **input** vector was presented to the network and the output was computed. If the output indicated...

...series was nonstationary, the series was differenced, the new ACFs were computed, and the new **input** was presented. Hoff states that this process should be continued until "... you are able to...

...only five of the series were nonstationary. The five nonstationary series were differenced and again **input** to the network. The **results** indicated that with the single differencing of the five nonstationary series, all 36 time series...

...2 neural network was used to identify appropriate stationary Box-Jenkins models. As described earlier, **input** vectors consisting of the first six ACF elements, the first six PACF elements, and a six-position target **vector** were constructed from the stationary **series** indicated by the application of the Stage 1 neural network. A dummy target vector of 100000 was used for all **inputs**. The output of the Stage 2 neural network indicated the Box-Jenkins model mappings.

To...

...of the neural network Box-Jenkins model recommendations, each of the 36 time series was **input** to the Forecast Pro expert system. After obtaining the expert system model recommendations, the Execustat statistical package was used to forecast all 36 **series** using the **model** recommendations **resulting** from both the expert system and those indicated by the neural networks. Table 6 summarizes the **results** with respect to the criteria outlined earlier.

NUMBER OF IDENTICAL MODEL RECOMMENDATIONS: 12  
INDICATIONS OF...

...LOWEST VALIDATION PERIOD RMSE            14            10  
TABLE 6.

Comparison of  
forecast data for  
similar length **series**.

Comparison of the two **model** identification systems revealed some interesting **results**. First, note from Table 6 that the expert system and the neural networks provided identical...

...recommended model's fit to the analysis period data showed that, among the 24 time **series** with differing recommended **models**, the neural network model forecasts produced lower RMSEs, and thus, better fits for 17 of...

...better validation period RMSEs, the expert system RMSEs were less than 10% lower than those **resulting** from the neural network models. Thus, in terms of the ability to satisfactorily forecast, the...

...model identification systems had virtually equal performance.

In summary, the analysis of 36 actual time **series** indicated that the Box-Jenkins **models** recommended by the neural network system provided better fits to the analysis period data used...how well the networks would generalize from the learned patterns to identify correctly Box-Jenkins **models** for **series** of greater length. To test the robustness of the neural networks in identifying appropriate Box-Jenkins **models** for actual **series** of greater lengths, 30 actual time series were taken from various sources. Fifteen of the...

...the 36 described in the previous section. The ACFs were computed in Execustat and the **input** vectors were constructed. The 30 **input** vectors were then presented to the Stage 1 neural network. The network outputs indicated that...

...to compute the ACFs and PACFs needed for the creation of the Stage 2 network **input vectors**. Presentation of the stationary **series input vectors** yielded the Box-Jenkins model recommendations.

After obtaining the Stage 2 neural network **results**, the 30 series were **input** to the Forecast Pro expert system and the model recommendations were recorded. Both the expert...

...network model recommendations were forecast using Execustat. Table 7 summarizes a comparison of the forecasts **resulting** from the two systems.

NUMBER OF IDENTICAL MODEL RECOMMENDATIONS: 17

INDICATIONS OF POOR MODEL QUALITY...

...7

TABLE 7.

Comparison of  
forecast data for  
series of length 100  
and 150.

The **results** in Table 7 show clearly that even though the neural networks were trained with data from series of length 50, they performed quite well in identifying appropriate Box-Jenkins **models** for longer time **series**. In fact, as shown in Table 7, identical models were recommended for 17 of the...

...validation period RMSEs revealed that the corresponding neural network models were twice within 10%. The **results** of this portion of the analysis showed that a neural network approach to Box-Jenkins...

...from series of length 50, the networks generalized in a way that often provided superior **results**, when compared to expert system **model** recommendations, for **series** that were twice and three times longer.

#### POWERFUL TOOLS

The problem of relating time **series** to appropriate Box-Jenkins ARIMA **models** has been approached in a variety of ways, including statistical model identification techniques and, more recently current literature to be powerful tools in relating sets of **inputs** to appropriate outputs in cases where no algorithmic relationship exists.

The aim of this study was to use artificial neural networks to identify appropriate Box-Jenkins **models** for time **series**. The **results** presented in this paper clearly indicated that a three-layer backpropagation artificial neural network is a superior method of identifying Box-Jenkins **models** for univariate nonseasonal time **series**. The application of neural networks produced **results** that were both accurate and robust. Based on our findings, it seems clear that an...  
...our work that will investigate the application of neural networks for solving the Box-Jenkins **model** identification problem for seasonal time **series** is both warranted and feasible.

#### SUGGESTED READING

Apolioni, B., G. Avanzini, N. Cesa-Bianchi, and...

...A Comparative Analysis." International journal of Forecasting 5: 209-215, 1989.

Vandaele, W. Applied Time **Series** and Box-Jenkins **Models**. New York: Academic Press Inc., 1983.

Zurada, J. M. Artificial Neural Systems. St. Paul, Minn...

19950600

31/3,K/26 (Item 22 from file: 484)  
DIALOG(R)File 484:Periodical Abs Plustext  
(c) 2005 ProQuest. All rts. reserv.

02740569 (USE FORMAT 7 OR 9 FOR FULLTEXT)

**Predicting the outcome of competition using experimental data: Maximum likelihood and Bayesian approaches**

Pascual, Miguel A; Kareiva, Peter

Ecology (GECO), v77 n2, p337-349

Mar 1996

ISSN: 0012-9658 JOURNAL CODE: GECO

DOCUMENT TYPE: Feature

LANGUAGE: English

RECORD TYPE: Fulltext; Abstract

WORD COUNT: 7539

LENGTH: Long (31+ col inches)

**Predicting the outcome of competition using experimental data: Maximum likelihood and Bayesian approaches**  
1996

ABSTRACT: A **statistical analysis** of Gause's experimental cultures of *Paramecium aurelia* and *P. caudatum* using analytical methods based...

TEXT:

... of population behaviors. It is thus important to consider this source of variability when drawing **conclusions** from experiments.

In this paper we present a **statistical** treatment of experimental **data** based on both maximum likelihood and Bayesian **statistics**. We show that, even when **analyzing** tightly controlled laboratory studies, significant levels of uncertainty about the nature of the competitive interaction...

...s competition experiments.

A review of Gause's (1934) experiments and simple Lotka-Volterra competition **models**

Gause performed **several** studies in which he monitored the growth of populations of species when each of the...

...competition model (Lotka 1924, Volterra 1926) has been used repeatedly to represent and analyze the **results** of Gause's experiments with *Paramecium* (Gause 1934, Leslie 1957, Roughgarden 1979). This model assumes ...

...of the application of the LV model.

The LV competition model generates four possible dynamic **outcomes**, depending on the values of the carrying capacities and the competition coefficients,

(omitted)

Thus, when the model is fitted to experimental **data**, we can hope to **predict** the dynamic **outcome** based on the values of the competition coefficients and the K's.

Methods

When confronted...it is not based on a precise definition of the error structure of the available **data**. Performing **statistical** tests or constructing confidence intervals with parametric methods requires clear assumptions about the distributional properties...

...structures for the variability in the observations and derive the corresponding probabilistic models; we then **search** for those parameter estimates that best represent the data and derive measures of uncertainty about...



...was preferred over a simple discrete time model because it allowed us to frame our **results** in the classical terms of the four different qualitative **outcomes** for two-species competition that depend on the alpha sub ij 's and K sub...

...the system are deterministic, but the population size at each time interval is known with **error**. Fitting the **model** consists of generating **deterministic trajectories**, contrasting the **trajectories** to the actual data, and selecting those parameter values that minimize the discrepancies between the...

...growing literature, especially in fisheries, on the relative merits of fitting process and observation error **models** to time **series** of population data. Failure to incorporate observation errors in the estimation **results** in biased parameter estimates (Walters and Ludwig 1981, Polachek et al. 1993). This is called the "error in variables" in the statistical literature (Kendall and Stewart 1973), and **results** from having sampling error in the independent variable. In linear regression problems, error in variables...error fit is also affected by the so-called time series bias (Walters 1985), which **results** from the fact that when dealing with time series, the value that the "independent" variable takes is not independent of the magnitude of the process error earlier in time. This **results** in an "unbalanced" sample of population behaviors, and a biased depiction of the average population...

...each error structure assumption in Gause's data, we obtained the ML estimates by numerical **search** for those values of theta at which the total likelihood was the highest. Unlike the...

...of residuals; for both process and observation fit, we inspected the residuals over time in **search** of deviations from the independence assumption, calculated the autocorrelation coefficients for individual experiments, and calculated...that subset) can be obtained using the likelihood profile method (Venzon and Moolgavkar 1988), by **searching** for those values of mu satisfying:

(omitted)

where (omitted) is the likelihood estimated at the...

...estimates and the joint limits for alpha sub 12 and alpha sub 21 by numerical **search** of those values that satisfied Eq. 8.

Bayesian estimation of parameters and density regions

We...

...previous section (also called frequentist), the parameters (B), regarded as the central objective of the **statistical analysis**, are unknown constants. Estimators for these parameters, which are functions of the data, are selected...

...the parameters before obtaining the data (Box and Tiao 1973). Given the prior distribution, the **data** from our experiment, and a **probability** model for the **data** (omitted) in Eq. 5

, it is now possible to calculate the posterior probability density, representing the probability of the parameters at the **conclusion** of the experiment (omitted). Bayes' theorem allows us to combine the prior distribution and the data to form the posterior distribution:

(omitted)

The term in the denominator is the total **probability** of the **data** and provides a normalizing constant to ensure that the posterior density integrates to 1. Because...

...provide a straightforward way to incorporate auxiliary information in

the estimation process. These two approaches **result** in what are called objective priors, because they have a direct frequency interpretation and are...Metropolis-Hastings algorithm (Hastings 1970), with 50000 simulations.

As an alternative way to summarize the **results** we also constructed intervals in the posterior densities. The intervals derived are called highest posterior...

...initial population densities at the start of the experiment. Furthermore, in some cases, the qualitative **outcome** of competition depends completely on the initial conditions (fourth case in Eq. 2). Finally, since...

...three different starting conditions. For each of the two experimental designs we generated 200 data **sets** by **simulating** deterministic trajectories using the MLEs (Eq. 3 with  $w's = 0$ ), but added a lognormal...

...chosen was used to quantify the relative merits of our two different experimental designs. The **probabilities** derived from this **analysis** represent how likely we are to detect the true dynamics of the system from the point estimates when different experimental designs are used.

### Results

The fit of the LV model to Gause 's data

Fig. 1 shows the best...scenario for Eq. 2 based on MLEs in Table 2), we arrive at two different **conclusions** depending on the estimates we use. Point estimates assuming observation error indicate that in Gause...

...characterized by a singular behavior, but by a probability distribution of alternative behaviors, and equilibrium **results** derived from Eq. 2 should not really be applied to interpretations of estimated model parameters...

...much larger populations of *P. caudatum* than of *P. aurelia* in the mixed culture; this **results** in a much better appraisal of *P. caudatum*'s per capita effect on *P. aurelia*...

...estimated by the process error method is much wider.

Fig. 3 shows the posterior densities **resulting** from the Bayesian analysis of Gause's data. Although the interpretations of Bayesian high posterior densities and confidence intervals using a likelihood approach are different, **results** from both methods parallel each other. For instance, when combining noninformative priors with normal or...

...joint confidence interval in Fig. 2. The Bayesian posteriors thus lead to the same general **conclusions** about competition between *Paramecium* species: there are significant mutual competition effects, with a possible larger...

...experiments showed positive significant autocorrelation at lag 1 and 2, suggesting that there may be **systematic** deviations from the **deterministic trajectories**. The **process error** assumption appears to take care of this correlation, and no significant autocorrelations in the residuals...

...that the system conforms to alternative scenarios. The main uncertainty comes from distinguishing between the **outcome** of competition depending on the initial conditions (57%) vs. *P. caudatum* always winning regardless of ...

...common alternative scenario selected was that in which *P. caudatum* always wins (20%). When experimental **sets** were **simulated** with contrasting initial numbers for both species in the mixed experiments, a

perfect evaluation of...s data shows that even well-controlled laboratory experiments can entail large uncertainty about the **outcome** of species competition. The analysis of point estimates can be highly misleading, and different statistical techniques can **result** in fundamentally different **conclusions**. Such uncertainties suggest that we need to be very careful about how we conduct our statistical inference.

While the validity of Bayes' **theorem** to reverse conditional **probabilities** is beyond question, other issues about its application to statistical inference have been extremely contentious...

...estimation process and, by choosing alternative prior distributions, authors enable readers to form their own **conclusion** (Edwards et al. 1963). Also, it is always possible to use noninformative priors that allow

...is important to note that the mechanics of the methods we used for fitting dynamic **models** to time **series** are very general. One may use literally any dynamic model that predicts changes in densities...

...s experiment documents how significant assumptions about error structure can be--we predict different qualitative **outcomes** for competition between Paramecium depending on whether we assume process error or observation error. In...

...be greatly increased by setting mixed cultures with contrasting initial conditions, and fitting a competition **model** to the whole time **series**. It will sometimes be more economical to manipulate initial conditions than to run experiments for...

...developed using Gause's Paramecium).

Finally, we found that although error assumptions profoundly influenced our **conclusions** about the qualitative **outcome** of Paramecium competition, the adoption of a Bayesian or maximum likelihood approach did not make much of a difference. In particular, both the Bayesian and likelihood approach support the **conclusion** that there are significant mutual competition effects, with a possible larger effect of *P. aurelia*... such as that provided by Table 3 would probably be the most useful distillation of **conclusions**. We think a Bayesian approach has general value for applied questions such as: will a...population dynamics. Biological Journal of the Linnean Society 32:271-279.

Berger, J. O. 1985. **Statistical decision theory** and Bayesian **analysis**. Springer-Verlag. New York. New York. USA.

Box, G. E. P., and G. C. Tiao. 1973. Bayesian inference in **statistical analysis**. Addison-Wesley, Boston, Massachusetts. USA.

Boyce, W. E., and R. C. DiPrima. 1986. Elementary differential...

...USA.

Carpenter. S. R., K. L. Cottingham, and C. A. Stow. 1994. Fitting predator-prey **models** to time **series** with observation errors. Ecology 75:1254-1264.

Clutton-Brock. M. 1967. Likelihood distributions for estimating...

...choice, dynamics and uncertainty. Chapman and Hall, New York, New York, USA.

Jeffreys, H. 1961. **Theory of probability**. Oxford University Press, London, England.

Kendall, M. G., and A. Stewart. 1973. The advanced **theory of statistics**. Haefner, New York, New York, USA.

Leslie, P. H. 1957. An analysis of the daa...

DESCRIPTORS: **Statistical analysis** ;

59/3,K/134 (Item 42 from file: 349)  
DIALOG(R) File 349:PCT FULLTEXT  
(c) 2005 WIPO/Univentio. All rts. reserv.

00409232 \*\*Image available\*\*  
MODEL -**BASED** FAULT DETECTION SYSTEM FOR ELECTRIC MOTORS  
SYSTEME DE DETECTION DE PANNES A L'AIDE D'UN MODELE POUR MOTEURS  
**ELECTRIQUES**

Patent Applicant/Assignee:

ARCELIK A S,  
DUYAR Ahmet,  
OSMAN Tugurul Durakbasa,  
EVREN Albas,  
SERAFETTINOGLU A Hakan,

Inventor(s):

DUYAR Ahmet,  
OSMAN Tugurul Durakbasa,  
EVREN Albas,  
SERAFETTINOGLU A Hakan,

Patent and Priority Information (Country, Number, Date):

Patent: WO 9749977 A1 19971231  
Application: WO 97TR8 19970620 (PCT/WO TR9700008)  
Priority Application: TR 96527 19960624

Designated States:

(Protection type is "patent" unless otherwise stated - for applications prior to 2004)

AL AM AT AU AZ BA BB BG BR BY CA CH CN CU CZ DE DK EE ES FI GB GE HU IL  
IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MD MG MK MN MW MX NO NZ PL PT  
RO RU SD SE SG SI SK TJ TM TR TT UA UG US UZ VN GH KE LS MW SD SZ UG ZW  
AM AZ BY KG KZ MD RU TJ TM AT BE CH DE DK ES FI FR GB GR IE IT LU MC NL  
PT SE BF BJ CF CG CI CM GA GN ML MR NE SN TD TG

Publication Language: English

Fulltext Word Count: 9385

MODEL -**BASED** FAULT DETECTION SYSTEM FOR ELECTRIC MOTORS  
SYSTEME DE DETECTION DE PANNES A L'AIDE D'UN MODELE POUR MOTEURS  
**ELECTRIQUES**

Fulltext Availability:

Detailed Description  
Claims

English Abstract

the present invention relates to a **model** -based **fault** detection system and method for monitoring and predicting maintenance requirements of electric motors. Since the...

Detailed Description

**MODEL** -**BASED** **FAULT** DETECTION SYSTEM FOR ELECTRIC MOTORS  
BACKGROUND OF THE INVENTION  
The present Invention relates to electric...

...to the output signals of complex systems with the output signal obtained from a mathematical **model** of the **fault** free system. The comparison of these signals is ified in terms of a "residual" which...

...accurate, real-time models for i-nultivariable systems, that is, systems with more than one **inputs** and/or one outputs.

If the **model** of the system is not accurate, the residuals will contain modeling errors that are very...

...actual faults.

Another shortcoming of such FDD methods relates to the difficulty in generating a **data** base for **statistical** testing of residuals to classify faults. Developing such a database requires a priori information about...

...data base must also meet the specific requirements of a particular FDD scheme.

Since, mechanical **faults** are the **result** of vibration, detection and analysis of .5 vibration is a common element of many prior...

...mounting and positioning, signature detection may be corrupted by background vibration and variation in operating **conditions** such as running speed, **input** voltage and motor loading.

It will be appreciated that the likelihood of erroneous indication of...

...a fault exists.

#### SUMMARY OF THE INVENTION

1.5 The present invention relates to a **model** based **fault** detection system and method for monitoring and predicting maintenance requirements of electric motors and more...

It may be selected to compensate for modeling errors and noise or other **perturbations** that may **result** in a non-zero residual.

When a motor component degrades such that the motor is...

...of the likely failed or failing component.

In another embodiment of the present invention, a **system** for **detecting** and diagnosing mechanical **faults** of fractional horsepower electric motors is disclosed.

Rather than developing an extensive database to **correlate faults** with the measured signals, the present embodiment incorporates a mathematical model of a fault... Tektronix 503b ac/dc current probe amplifier). Signals from sensors 34 and 35 are also **conditioned** by board 37 and **input** to computer 42. Computer 42 records sensor data in its memory (not shown).

Computer 42 implements a **fault** detection and diagnostic **model** of an ideal motor which is also stored in memory. In the preferred embodiment, the... known in the art and as described in the above-referenced I-). apci-.

The **model** based **fault** detection and diagnostic scheme of the present invention (ICSGIA) is Li IULLII free MOLOr With...

...appropriate dimensions.

In the fractional horsepower electric motor, by way of example, the experimental **model** uses **input** voltage, current and speed measurements.

Ill Fl, (.Ture 3, a plot of input voltage 38...

...up motor 10 is shown. In the preferred embodiment, input voltage 38 is a step **input** and is represented in the experimental **model** as a row vector containing the measured voltage. Figure 4 shows the experimentally determined current...and the modeling errors, the residual vector  $r(k)$  equal to (lie zero vector under **Fault free conditions** . A nonzero value of the residual vector indicates the existence of faults. When noise and...

...comparing the residual magnitudes with selected threshold values. Using the observed distribution of residuals under **fault free conditions** , the threshold values are determined by selecting a level of confidence (within three standard deviations...is equal to the zero vector and correspondingly that the motor is operating under **fault free conditions** . If comparator 48 determines that the residual vector  $r(k)$  has a nonzero value, one...Table 2. As will be noted, the inductance,  $L$ , of faulty motor 10 exceeds the **corresponding** inductance parameter predicted by **model** 44 by more than one standard deviation while all other parameters are less than the...

...speed,  $w$ , of motor 10 at step 78 and compares these 1.)arameters with the **corresponding** parameters of **model** 44 (see also Table 2).

based on the comparison, system 30 is able to classify...reliable technique or measurement to identify fault free motors, first a method to obtain the **model** of typical **fault free** motors (the "base **model** ") is developed.

A more detailed explanation of the MQM. method is depicted in Figures 8A ...deviations are obtained and stored in the designated M data f-11C.

'Hic hiise motor **model** is developed from a **group** of motors known to include mostly 11IL111 free motors, step 96. In one preferred embodiment ...

...evaluated for obvious modeling and threshold errors, steps 102

I J sing the base motor **model** developed from the **group** , each of the motors in [lie group is then tested against the experimental base motor ...

...of motors remaining in the group.

eliminating all motors having outputs outside of tolerance values **set** by the experimental **model** , it is possible to further yet refine the experimental **model** by evaluating modeling **errors** , the mean and standard deviations of the group, step 114 'I the group contains...

...Various options may be presented to the user. By way of example, the base motor **model** may **correspond** to a universal, shaded pole induction motor, synchronous motor or any other fractional horsepower electric... exceed the pre-selected threshold values, the motor being tested is defined as having a **fault** .

@pecially, once the base motor **model** is selected, the user **inputs**

the necessary I O parameters in performing the "Quality Assurance Test" at steps 130-134...of the faulty motor are evaluated by the MQM algorithm and compared with the **corresponding** parameters of the base motor **model**, steps 156 The result of this comparison is used to classify the motor fault and...

...are evaluated. These parameters are compared with the physical parameters of the base motor **model** using the **fault** classification threshold values (see Table 4). For the universal electric motor, the physical parameters are...

...effect of changes in each parameter on the outputs of the motor. In such a **situation** the **model** may have **multiple**, but small, **faults** which may cumulate to cause the model outputs to exceed the threshold values. However, since...

...repair shops to diagnose faults and preventative maintenance purposes. In such an application base motor **models** for **several** electric motors, varying by size and ...customized model is developed when it is known that the motor is running under **fault free conditions**. In contrast, the MQM develops a base model that encompasses the variations normally associated with...

...population. Accordingly, it is possible for a marginally operating motor to pass the test thresholds **set** in the MQM **model** but it is unlikely that continued degradation will go undetected by the MCM since the...  
...motor is dependent on the requirement imposed by the application.

One may appreciate that the **input** applied to **model** 44 may not be as "rich" of an input signal as could be applied during...

#### Claim

... L1C1- means coupled to said sensors for determining, when said motor is operating in a **fault free condition**, an ideal residual of zero, for selecting a threshold tolerance level different than zero...

...on said display device a message indicative of whether said motor is operating under **fault free conditions** in the case where said **plurality** of residuals is less than the threshold tolerance or whether said motor is operating with...

...whether said motor is operating without  
if detected fault;  
correcting said changes to a mechanical **fault** in the **event** said motor is operating with a detected fault and communicating the existence of said fault...model having a threshold tolerance based on two standard deviations about the mean of the  
**group model** ;  
testing each motor from said **group** by re-measuring said plurality of operating signals and removing said motor from said group...

...The method of claim 19 further comprising the steps of correlating said changes to a mechanical **fault** in the **event** said motor is operating with a detected fault and communicating the existence of...

... $\frac{dw}{dt} + fw = k_1 i^2 + M$   
where  $k_1$  and  $k_2$  are motor constants.

22 A **model** based **fault** detection and diagnosis system for detecting faults in a group of motors and developing diagnostic information for correction of said faults

comprising:

means for generating a system **model** representing a mean of **fault** free motors

according to the following equations:

$x(k+1) = Ax(k) + Bu(k)$  and  $y(k) = Cx(k)$

where the A, B and C represent parameters of said system **model** ; and means for determining system **faults** by measuring parameters of said motors with comparing Nvith said system model.

23 'I@liesystemofcialiii22whereinsaideterminingmeanscomprisescomputer... display device a message M indicative of whether each of'said motors are operating under **fault** free **conditions** in Ilic case where the difference between said equations representing each of said motoi-s...